Development of machine vision algorithms for radiomics and advanced particle therapy

Department of Electrical and Information Engineering

ELECTRICAL AND INFORMATION ENGINEERING Ph.D. Program

SSD: PHYS-01/A – ING-INF-07

Final Dissertation

Development of machine vision algorithms for radiomics and advanced particle therapy

by

Ferrara Nicola

Coordinator of Ph.D. Program:
Prof. Mario Carpentieri

Supervisors:
Prof.ssa Gabriella Pugliese

Prof. Filippo Attivissimo

Course n°38, 01/01/2023-31/12/2025

Department of Electrical and Information Engineering

ELECTRICAL AND INFORMATION ENGINEERING Ph.D. Program

SSD: PHYS-01/A – ING-INF-07

Final Dissertation

Development of machine vision algorithms for radiomics and advanced particle therapy

by

Ferrara Nicola

Nicola Ferrara

Coordinator of Ph.D. Program:

Prof. Mario Carpentieri

Mario Carpentieri

Supervisors:

Prof.ssa Gabriella Pugliese

Gabriella Pugliese

Prof. Filippo Attivissimo

Filippo Attivissimo

Referees:

Prof. Giacomo Borghi

Prof. Andrea Cataldo

Course n°38, 01/01/2023-31/12/2025

Table of Contents

Introduction

This thesis investigates the use of Deep Learning-based approaches, to address two primary computational imaging challenges associated with cancer treatments.

First, a Deep Learning (DL) framework is implemented for semantic segmentation of pancreatic tumors from medical imaging datasets. The use of DL for automatic segmentation of pancreatic tumors in medical imaging modalities such as Computed Tomography (CT) is investigated. Precise segmentation is essential not only for diagnosis and staging but also for defining target volumes in radiation therapy [1–3].
One of the most probable cause of death nowadays is cancer. One of the most aggressive is the pancreatic cancer, that remains one of lethal forms of cancer, characterized by late diagnosis, limited therapeutic options, and poor prognosis [4]. Pancreatic Ductal Adenocarcinoma Cancer (PDAC) remains one of the most deadly cancers, with a five-year survival rate below 10 percent, primarily due to asymptomatic early stages and rapid metastatic progression. Precise localization and delineation of the tumor are critical for diagnosis, staging, and the planning of advanced therapies. Despite advancements in medical imaging and treatment, early and precise detection of pancreatic tumors continues to present a formidable challenge. One of the architectures used is the U-Net's encoder-decoder structure, that enhanced with skip connections, enables precise boundary delineation, which is vital for accurate tumor volume estimation and radiation treatment planning [5, 6].

Second, in this thesis is proposed a DL based method for reconstructing and enhancing Single Photon Emission Computed Tomography (SPECT) images, which are used to monitor the in vivo distribution of boronated compounds for an innovative therapeutic techniques, the Boron Neutron Capture Therapy (BNCT), that offers a promising direction for targeted cancer treatment[7, 8]. It is important to state clearly that while full *in vivo*

dose reconstruction in BNCT using a tomography setup based on gamma cameras is the ultimate long-term clinical goal, the scope achieved within this framework focuses on foundational feasibility, specifically the reconstruction of boron spatial distribution in controlled simulated and experimental phantom setups. BNCT relies heavily on accurate imaging for patient selection, treatment planning, and post-therapy evaluation, SPECT playing a central role in visualizing boron distribution and monitoring therapeutic efficacy.

BNCT relies heavily on accurate imaging for patient selection, treatment planning, and post-therapy evaluation, SPECT playing a central role in visualizing boron distribution and monitoring therapeutic efficacy [9].
It was examined the role of U-Net-based models in the reconstruction and enhancement of low-dose or under-sampled SPECT images. High-quality SPECT reconstruction is crucial in BNCT to accurately localize boron uptake and assess treatment distribution, while minimizing radiation exposure to patients.
Traditional iterative reconstruction algorithms, such as Ordered Subsets Expectation Maximization (OSEM), are computationally intensive and sensitive to noise, particularly in low-count imaging scenarios [10, 11]. Deep models can learn spatial priors and statistical correlations to recover high-fidelity reconstructions from sparse or noisy projections, offering improved resolution and noise suppression.
By integrating these two approaches, this work contributes toward a robust deep learning pipeline capable of automating key imaging tasks in the BNCT workflow. The segmentation model facilitates accurate tumor targeting, while the reconstruction model enhances the quality and interpretability of SPECT images critical for therapy planning and response assessment. The proposed models are evaluated on available acquired datasets made by simulation and measurements, using standard metrics such Structural Similarity Index Measure (SSIM), and Peak Signal-to-Noise Ratio (PSNR), with emphasis on generalizability and computational efficiency.
By exploiting the powerful feature extraction and spatial learning capabilities of DL models, this research aims to improve the reliability and automation of pancreatic cancer segmentation and SPECT image reconstruction. The combined approach contributes toward more accurate diagnostics and effective treatment planning in BNCT, ultimately enhancing patient outcomes. By exploiting the powerful feature extraction and spatial learning capabilities of DL models, this research aims to improve the reliability and automation of pancreatic cancer segmentation and SPECT image reconstruction. The combined approach contributes toward more accurate diagnostics and effective treatment

planning in BNCT, ultimately enhancing patient outcomes. To make this work accessible to a broader audience, it is important to contextualize the core methodologies. Image reconstruction is a fundamental process in medical imaging, particularly in modalities such as Computed Tomography (CT), which uses X-rays to generate structural maps of tissue density, and emission tomography (like PET and SPECT), which detects functional and metabolic processes using radiotracers. While traditional reconstruction relies on analytical or iterative mathematical models, Deep Learning (DL) has recently revolutionized this field. DL models, particularly Convolutional Neural Networks (CNNs), are computational architectures inspired by the human brain that learn to extract hierarchical features from large datasets. By iteratively adjusting their internal parameters during a training phase, these models can identify complex spatial patterns, making them exceptionally suited for tasks such as automated tumor segmentation and the enhancement of noisy or sparse reconstruction data.

Limitations Despite the promising results, this study presents certain limitations. First, the deep learning models were trained and evaluated on specific, relatively limited datasets (e.g., MSD and CPTAC-PDA for segmentation, and simulated FLUKA data for SPECT). This may limit the generalization of the models to unseen data acquired from different clinical scanners with varying noise profiles and protocols. Second, the 2-step segmentation pipeline, while improving accuracy, introduces an accumulation of computational overhead, and localization errors in the first step (e.g., bounding box misalignment) inevitably propagate to the fine-segmentation stage. Finally, regarding the BNCT-SPECT reconstruction, the reliance on simulated background distributions and static detector geometries necessitates further validation in dynamic, real-world clinical environments where background interference is highly variable.

Code and Data Availability The datasets utilized in the segmentation study (Medical Segmentation Decathlon and CPTAC-PDA) are publicly available through their respective repositories. The proprietary Geant4 and FLUKA simulation data, as well as the experimental measurements from the LENA facility, are available from the corresponding authors upon reasonable request. The code and scripts used for model training, preprocessing, and tomographic reconstruction (implemented in TensorFlow and PyTomography) are available upon reasonable request to allow full reproduction of the experiments.

Chapter 1 This chapter provides a rigorous mathematical analysis of Computed Tomography (CT). It explores the physics of the Radon transform and the projection-slice theorem, which constitute the theoretical basis of CT imaging. The discussion categorizes reconstruction algorithms into two main families: analytic methods, such as Filtered Back Projection (FBP), and Iterative Reconstruction Techniques (IRT), detailing their respective mathematical formulations, convergence properties, and computational requirements.

Chapter 2 This section reviews the state-of-the-art DL models employed in medical imaging. For tomographic reconstruction, the chapter analyzes Convolutional Neural Networks (CNNs), Autoencoders (AE), and Variational Autoencoders (VAE), discussing their ability to resolve inverse problems. Regarding segmentation, the focus shifts to architectures such as the U-Net and Vision Transformers (ViT). The chapter concludes with a survey of current applications of these models in clinical settings.

Chapter 3 This chapter presents a specific case study on the automatic detection and segmentation of pancreatic lesions. Using the Medical Segmentation Decathlon (MSD) dataset containing 281 labeled CT scans , the study implements a rigorous preprocessing pipeline involving windowing (cut of -100 to 240 HU) and normalization. Data augmentation techniques, such as shifting, were applied to enhance model robustness. The core of the analysis compares a 1-step pipeline against a 2-step pipeline. The 2-step approach first utilizes a YOLOv11 CNN for tumor localization , followed by a U-Net architecture to segment the specific tumor region. Results demonstrate that the 2-step pipeline significantly outperforms the segmentation without localization, achieving a DSC of 58% compared to 45.50% for the 1-step approach.

Chapter 4 The final chapter investigates SPECT applied to BNCT. It details the construction of simulations for an experimental facility and the measurement protocols used to reconstruct the boron dose within test vials. The chapter compares two reconstruction methodologies: the classical Ordered Subset Expectation Maximization (OSEM) algorithm and a novel DL reconstruction algorithm based on U-Net, trained on the simulated dataset. The performance of these approaches is evaluated and discussed in terms of reconstruction efficiency and quantitative metrics, specifically Intersection over Union (IoU), SSIM and PSNR.

Conclusions The conclusions report the successfully demonstration of the efficacy of DL algorithms in addressing complex segmentation and particle therapy challenges. The results of the segmentation of pancreas prove that the DL can actually help diagnosis of the most deadly human cancer. Overall, the integration of machine vision algorithms into the medical physics workflow presents a pathway toward more personalized and precise cancer treatments, bridging the gap between advanced computational methods and clinical application. The results in the SPECT study prove that the application of DL can reduce the amount of time for a dose monitor system.

Chapter 1

Tomography methods for medical imaging

Medical imaging techniques such as Computed Tomography (CT), Positron Emission Tomography (PET), and Magnetic Resonance Imaging (MRI) are built upon solving what is known as the inverse problem. This mathematical framework is essential for reconstructing the internal structure of an object, such as a human body, from a limited set of external measurements. Unlike the direct problem which involves computing the expected measurement given a known object the inverse problem seeks to infer unknown structures from observed data. In the context of tomography, this entails deducing spatial distributions of physical properties (e.g., attenuation coefficients or emission intensities) from a finite number of projections [12].
The inverse problem is inherently ill-posed: it may lack a unique solution, or small variations in the data can cause large variations in the output. This characteristic makes it particularly challenging, requiring sophisticated mathematical tools, robust numerical methods, and often, regularization strategies to ensure stability and accuracy. The significance of solving the inverse problem efficiently and accurately cannot be overstated, as it underpins the diagnostic capability of modern imaging systems.

1.1 Tomography

Tomography is a medical imaging technique used to create detailed cross-sectional images of the body. It works by taking multiple X-ray, magnetic, or ultrasound measurements from different angles and then using computer processing to reconstruct slices of internal

structures, such as organs, bones, and tissues. The most common forms of medical tomography include CT, MRI, PET and SPECT. These methods help doctors diagnose and monitor a wide range of conditions, from tumors and infections to vascular diseases and injuries, by providing a clearer view of areas that are difficult to assess with traditional X-rays. In nuclear medicine, the primary tomographic modalities are SPECT and PET. Both methods involve the administration of radiopharmaceuticals that emit gamma rays, which are detected by specialized cameras. The resulting data is reconstructed into three-dimensional images that reflect physiological and metabolic processes rather than purely anatomical structures. This makes nuclear medicine tomography particularly valuable for detecting functional abnormalities at early stages, such as in oncology for tumor detection, in cardiology for assessing myocardial perfusion, and in neurology for evaluating brain disorders like epilepsy and dementia [13, 14].

1.1.1 Inverse problem

Tomography is fundamentally a process of reconstruction of turning indirect measurements into meaningful images of the interior of an object or body. The heart of this process is the inverse problem: deducing the internal structure or properties of an object from data collected on its exterior or along its boundary [15]. In the context of medical imaging, this means inferring tissue density, radiotracer concentration, or other physical properties from the signals recorded by imaging detectors. When a patient undergoes a CT scan [16] or a PET study, detectors do not directly record an image. Instead, they collect projections aggregated data that represent the sum or integral of a quantity (like X-ray attenuation or gamma photon emissions) along a line or path through the body. The challenge is then to reconstruct the original distribution from these projections, a process mathematically represented as inverting a transform, such as the Radon transform in CT or the attenuated Radon transform in SPECT [17–19].

To understand the inverse problem, it is useful to contrast it with the forward problem. In tomography, the forward problem involves predicting what the measured projections would be, given a known internal distribution. This is usually well-posed and stable if the interior structure is known, one can simulate the measurements accurately using physical models of radiation transport, attenuation, and detector response.

The inverse problem, in contrast, may not have a unique solution, the solution may not depend continuously on the data, or a solution might not even exist without further constraints. This property arises because real-world measurements are noisy, incomplete, and sometimes inconsistent due to physical limitations or patient movement. As a result,

direct inversion methods can lead to artifacts or instability in the reconstructed image [20–22].

1.1.2 Radon and Fourier Transform

The Radon Transform can maps a function in 2D space to a set of line integrals.
In CT, it represents the X-ray attenuation along different paths through the object.
The Fourier Transform can converts spatial domain information to frequency domain. It's crucial in image processing and reconstruction algorithms.
The Fourier Slice Theorem connects the Radon and Fourier transforms, stating that the 1D Fourier transform of a projection is equivalent to a slice of the 2D Fourier transform of the original image.
These transforms are essential in CT image reconstruction, allowing the conversion of projection data into cross-sectional images [23]. Understanding these transforms is crucial for developing and optimizing reconstruction algorithms in medical imaging, particularly in CT and SPECT modalities.
Computed Tomography (CT) is one of the most widely used medical imaging techniques, providing detailed cross-sectional images of internal anatomical structures. CT imaging relies on capturing multiple projections of an object from different angles and then reconstructing those projections into a coherent volumetric image. Central to this process are two mathematical operations: the Radon Transform and the Fourier Transform [24]. These tools form the foundation of most analytical reconstruction techniques, enabling the translation of projection data into spatial representations. This chapter explores how the Radon and Fourier Transforms are employed in CT image reconstruction, focusing on their mathematical formulation, practical implementation, and significance in modern medical imaging.

1.2 Filtered Back Projection

The Radon Transform is a mathematical operation that maps a two-dimensional function (representing an image) into its line integrals across various angles. In the context of CT, it represents the theoretical basis for the data collected during a scan. When X-rays pass through the body, detectors record the attenuation along multiple paths; these measurements correspond to line integrals of the tissue density. The mathematical definition is: let $f(x, y)$ represent the attenuation coefficient at position (x, y). The Radon

Transform $Rf(\theta, s)$ is given by:

$$Rf(\theta, s) = \int_{-\infty}^{\infty} f(s\cos\theta - t\sin\theta, s\sin\theta + t\cos\theta)dt. \tag{1.1}$$

In formula:

1. θ is the angle of projection, s is the distance from the origin along the projection line,
2. t parametrizes the integration path,
3. s parametrizes the direction orthogonal to the integration path.

Practical implications are that in CT images, this formulation explains how an image is "projected" into sinograms two-dimensional arrays representing the X-ray intensities measured at different angles θ and positions s, as shown in Fig. 1.1. Each row of a sinogram corresponds to one angle, and each column to a detector position. The Fig. 1.1 explains better that concepts, taking an illustrative density function of a simulated brain, and making the sinogram plot.

The goal of reconstruction is then to invert the Radon Transform to recover the original function $f(x, y)$, which represents the internal structure of the body.
The Fourier Slice Theorem (also known as the Central Slice Theorem) provides the critical link between the Radon and Fourier Transforms, enabling efficient reconstruction algorithms.
The theorem states that the 1D Fourier Transform of a projection at angle θ corresponds to a slice of the 2D Fourier Transform of the original image, taken at the same angle through the origin. By the formula, if

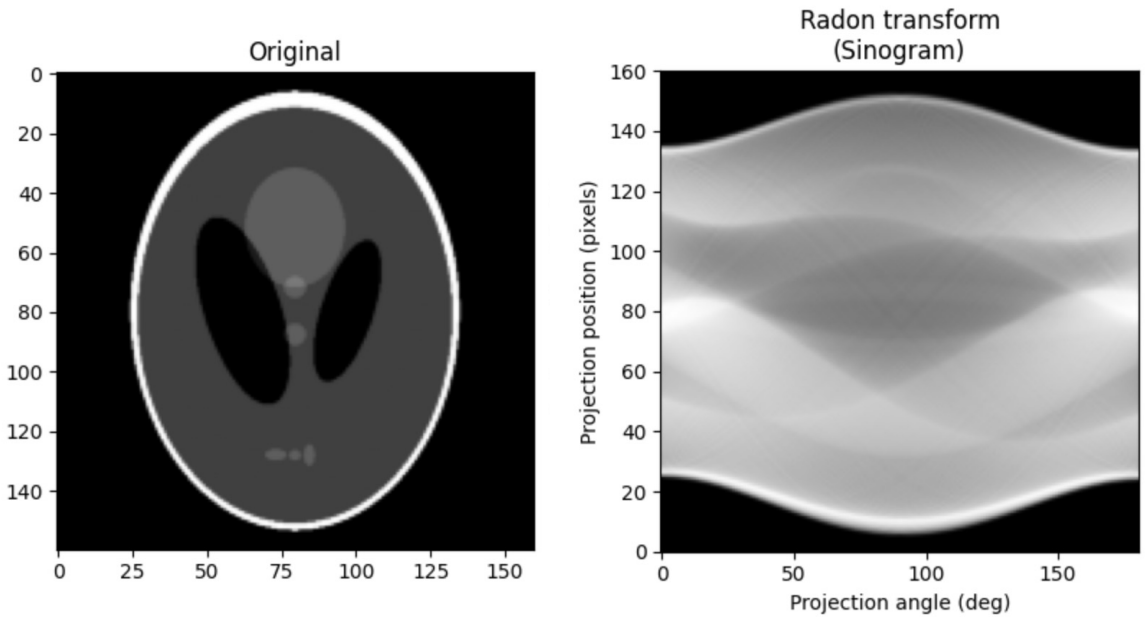
$$F(u, v) = \boldsymbol{F_{2d}}(f(x, y)) \tag{1.2}$$

is the fourier transform of the f in 2-dimensions, while

$$P_\theta(\omega) = \boldsymbol{F_{1d}}(R_\theta f(x, y)) \tag{1.3}$$

is the fourier transform in 1-dimension of the radon transform of f along θ-direction, the theorem states that

$$F(\omega, \theta) = F(u, v)|_{\omega,\theta} = P_\theta(\omega) \tag{1.4}$$



(a) (b)

Figure 1.1: (a) The Sheep-Logan image of a simulated brain. (b) The sinogram of the image: the projection position is measured in pixels, but is mathematically the s variable in Eq. 1.1. The projection angle is measured in deg and it is mathematically the θ variable in Eq. 1.1

.

This powerful relationship allows the use of Fourier-based techniques to reconstruct the image from its projections. Essentially, each 1D projection provides a radial line in the 2D frequency domain. By collecting enough projections from different angles, one can fill the 2D frequency space and then apply the inverse Fourier Transform to recover the image.

The Fourier Slice Theorem is the cornerstone of several analytical reconstruction algorithms, especially Filtered Back Projection (FBP), which is discussed next. It justifies why accurate reconstruction is possible with a complete set of projections, and why interpolation in frequency space is often necessary when data is discretely sampled.

The Filtered Back Projection algorithm is the most widely used analytical technique in CT image reconstruction. It is grounded in the theory of the Radon and Fourier Transforms and efficiently approximates the inverse Radon Transform. The algorithm follows these steps: Projection: Collect raw data (sinogram) from multiple angles at every step. Then apply a high-pass filter to each 1D projection in the frequency domain to compensate for the blurring that occurs in unfiltered back projection.Finally, spread the filtered projection data back across the image plane along the original paths of

integration.

Mathematically, this process can be summarized as [25]:

$$f(x,y) = \int_0^{\pi} \left(\frac{1}{2\pi} \int_{-\infty}^{+\infty} (h(w) * P_\theta(\omega)) e^{i\omega t} d\omega \right) \Bigg|_{t=x\cos\theta + y\sin\theta} d\theta \tag{1.5}$$

where

- $h(w)$ is the filter function,
- $*$ is the mathematical convolution operation

The filtering step relies on the 1D Fourier Transform of each projection: first, it transforms each projection to frequency domain.Second, it multiplies by a ramp filter to enhance high frequencies. Finally, it applies the inverse Fourier Transform to obtain the filtered projection.
This process highlights the utility of the Fourier Transform in isolating and correcting frequency components critical for image sharpness. The filters used are several: ramp, hann, hamming, cosine and shepp-logan. While ramp is used when it needs to reconstruc better the high frequency domain, the other are used to smoothing the edges, because they reduce the high frequency domain. In Fig. 1.2 some filters are shown.
The Radon and Fourier Transforms are not limited to basic 2D parallel-beam CT. They extend to a variety of geometries and advanced techniques in modern imaging.
In clinical practice, CT scanners use fan-beam or cone-beam geometries rather than parallel beams. These setups require modified Radon Transform models and reconstruction algorithms, but the foundational principles still hold. The Radon Transform is adapted to account for divergence, and back projection becomes a 3D operation.

1.3 Iterative methods

An approach that is not analytic, is the Iterative Reconstruction Technique (IRT). While FBP is fast and effective, it struggles with noisy or incomplete data. IRT do not explicitly use the Radon or Fourier Transforms, these methods often begin with analytical reconstructions or incorporate transform-based priors [26]. Examples include:

- Algebraic Reconstruction Technique (ART)

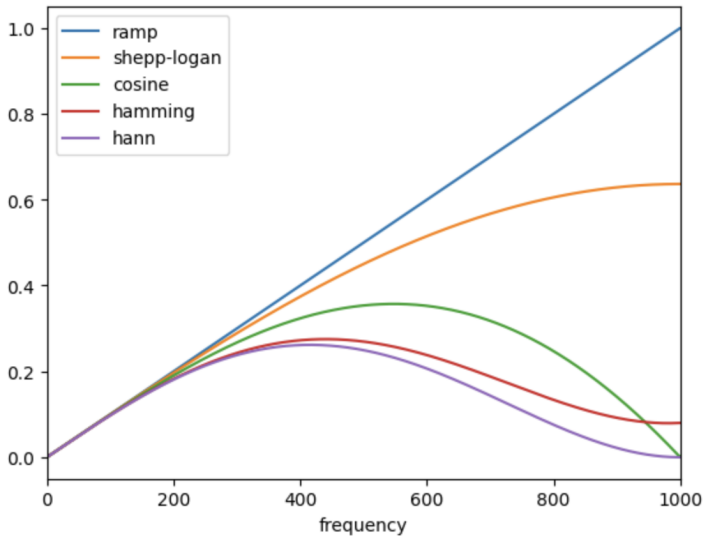


Figure 1.2: The filters used in the FBP. The ramp filter is frequently used to enhance the high frequencies. The shepp-logan to reduce high frequencies and the noise. The cosine, hamming and hann reduce noise but also enhance the blurring of the image.

- Simultaneous Iterative Reconstruction Technique (SIRT)
- Maximum Likelihood Expectation Maximization (MLEM)
- Ordered Subset Expectation Maximization (OSEM)

Modern approaches increasingly combine analytical methods with data-driven techniques. For example: DL Denoisers trained in the frequency domain, Learned filters replacing classical ramp filters, Neural networks mimicking FBP pipelines but with trainable components [27]. These methods leverage the foundational understanding of transforms while enhancing adaptability to real-world data variability [28].

The core logic of iterative reconstruction can be described as a cyclical optimization process, and Fig. 1.3 shows the algorithm that in general is used as starting point structure for all the methods described below. The goal is to find an image estimate $\hat{f}$ such that its forward projection matches the measured projection data g as closely as possible.

Let:

- f: The true tracer distribution (the image we want to recover).
- g: The measured projection data (sinogram).
- H: The system matrix (representing the probability that a photon emitted from voxel j is detected in bin i).
- $\hat{f}^{(k)}$: The image estimate at iteration k.

The relationship is generally modeled as $g = Hf + \text{noise}$. The noise is modeled randomly. The pipe-line of the process is the following. The first step consists of an initial guess: the algorithm initializes with a starting estimate, $\hat{f}^{(0)}$.

$$\hat{f}_j^{(0)} = \text{constant} \quad (\text{e.g., uniform value } > 0) \tag{1.6}$$

This serves as the baseline for the first iteration.The second step makes the forward projection (Simulation). In this step, the algorithm simulates the acquisition process using the current image estimate. It calculates the "expected" projections based on the current guess.

$$\hat{g}_i^{(k)} = \sum_j H_{ij} \hat{f}_j^{(k)} \tag{1.7}$$

Here, $\hat{g}^{(k)}$ represents the synthetic projection data at iteration k. In the third step follows the comparison with simulation and is performed the error calculation. The synthetic projections $\hat{g}^{(k)}$ are compared to the actual measured projections g. The nature of this comparison depends on the specific algorithm (e.g., difference for ART, ratio for MLEM). For Maximum Likelihood Expectation Maximization (MLEM), the ratio is calculated:

$$\text{Error Term}_i = \frac{g_i}{\hat{g}_i^{(k)}} \tag{1.8}$$

If the estimate is perfect, this ratio is 1. Deviations indicate where the image needs correction. The fourth step consists of the backprojection. The error term is backprojected from the projection space into the image space to update the estimate. This distributes the error among the voxels that contributed to it.
The MLEM update rule is:

$$\hat{f}_j^{(k+1)} = \hat{f}_j^{(k)} \cdot \frac{1}{\sum_i H_{ij}} \sum_i H_{ij} \left(\frac{g_i}{\sum_l H_{il} \hat{f}_l^{(k)}} \right) \tag{1.9}$$

Where:

- $\hat{f}_j^{(k)}$ is the current pixel value.
- The term in the parenthesis is the comparison (ratio) from Step 3.
- The summation $\sum_i H_{ij}(\dots)$ represents the backprojection of the error.
- $\frac{1}{\sum_i H_{ij}}$ is a normalization factor based on sensitivity.

Finally, the iteration is performed. Steps 2 through 4 are repeated. The updated image $\hat{f}^{(k+1)}$ becomes the current estimate for the next loop.

$$\hat{f}^{(0)} \to \hat{f}^{(1)} \to \cdots \to \hat{f}^{(n)} \tag{1.10}$$

The loop continues until a stopping criterion is met (e.g., k reaches a maximum number of iterations or the change in likelihood is negligible).

Algorithm 1 Generic Iterative Reconstruction Loop

1: Initialize $\hat{f}^{(0)}$
2: **for** $k = 0$ to $N_{iterations}$ **do**
3: **Forward Project:** Generate $\hat{g}$ from $\hat{f}^{(k)}$
4: **Compare:** Calculate error between $\hat{g}$ and measured g
5: **Backproject:** Map error back to image space
6: **Update:** Compute $\hat{f}^{(k+1)}$ using backprojected error
7: **end for**

1.3.1 Algebraic Reconstruction Technique

The Algebraic Reconstruction Technique (ART) is a family of iterative algorithms used in computed tomography (CT), emission tomography (like PET and SPECT) [30], and other imaging applications to reconstruct an image from its projections. ART is particularly useful when dealing with incomplete or noisy data, or when using sparse sampling. The problem could be mathematically summarized: the projections are $b \in \mathbb{R}^m$ in the measured projection data space, the *sinogram*. The image vector is the flatten image $x \in \mathbb{R}^n$. The matrix that describes the projection process is the *system matrix* $A \in \mathbb{R}^{m \times n}$.The formula to express the process is simply a linear system:

$$Ax = b \tag{1.11}$$

where the vector x represents the image to be recovered, b contains the measured projection data, and the matrix A encodes the imaging geometry, with each row a_i corresponding

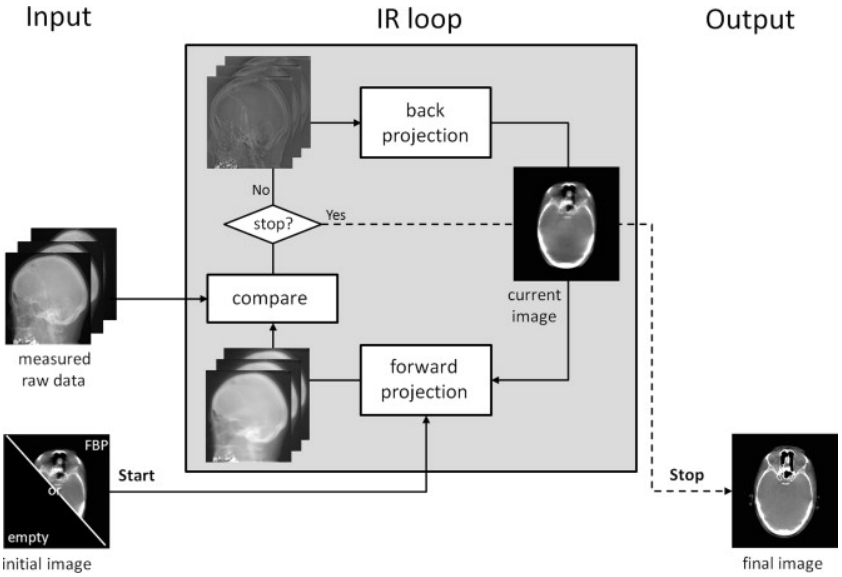


Figure 1.3: Explaining image for the Iterative Reconstruction Technique. The measured projections are compared with the initial forward projection of a sample image. Next iteration takes the error and map back to image space. Then computing the image by backprojection. Last make the forward for next comparison and the loop continues until comparison is good. [29].

to a single ray. ART updates the current estimate by successively projecting it onto the hyperplanes defined by the equations $a_i x = b_i$. The iterative update follows the Kaczmarz method and is given by

$$x^{(k+1)} = x^{(k)} + \lambda \frac{b_i - \langle a_i, x^{(k)} \rangle}{\|a_i\|^2} \, a_i^T \tag{1.12}$$

where λ is a relaxation parameter typically chosen in the range $0 < \lambda \leq 1$. At each step, the solution is corrected in the direction of a_i in proportion to the prediction error $b_i - \langle a_i, x^{(k)} \rangle$, ensuring that the final estimate progressively satisfies the measured projections. ART is effective when dealing with noisy or incomplete data, although it can be sensitive to the order of updates and to the choice of the relaxation parameter.

1.3.2 Simultaneous Iterative Reconstruction Technique

The Simultaneous Iterative Reconstruction Technique (SIRT) [31] is an iterative method for tomographic image reconstruction in which the entire image is updated simultaneously at each iteration based on the discrepancy between measured and simulated projections.

Given measured projection data p and system matrix W, SIRT proceeds with the following pipe-line. First, it undergoes the initialization. It is chosen an initial image

estimate $f^{(0)}$ (e.g. zeros or a filtered backprojection reconstruction). The second step makes the forward projection. It is computed the simulated projections

$$\hat{p}^{(k)} = W f^{(k)} \tag{1.13}$$

The third step consists in computing the projection error.

$$e^{(k)} = p - \hat{p}^{(k)} \tag{1.14}$$

The fourth step makes the backprojection of the error. It is computed the backprojection of the weighted error:

$$\Delta f^{(k)} = W^{\top} D \; e^{(k)} \tag{1.15}$$

where D is a diagonal matrix that normalizes each projection ray, commonly

$$D = \text{diag}\left(\frac{1}{\sum_j W_{ij}}\right) \tag{1.16}$$

The fifth step makes the normalization and the update of the image. Image is normalized by the number of rays through each pixel via a diagonal matrix C:

$$f^{(k+1)} = f^{(k)} + \lambda \; C \; \Delta f^{(k)} \tag{1.17}$$

where λ is a relaxation parameter, typically $0 < \lambda \leq 1$. Finally the steps 2 to 5 are repeated until a desired number of iterations or a convergence criterion is met.

Properties The algorithm shows several features: it produces smooth, stable reconstructions and is robust to noise and limited-angle data. The algorithm converges to a least-squares solution of $Wf = p$.It is typically slower than analytic methods such as filtered backprojection. It is very straightforward to incorporate regularization or constraints.

1.3.3 Maximum Likelihood Expectation Maximization

The Maximum Likelihood Expectation Maximization (MLEM) algorithm is an iterative method for tomographic image reconstruction derived from maximizing the Poisson likelihood of the measured projection data. It is widely used in emission tomography (PET/SPECT) and can also be applied to transmission tomography.

Forward Model Let p_i denote the measured projection for ray i, and let the system matrix W_{ij} describe the contribution of voxel j to ray i. The expected projection for ray i is

$$\hat{p}_i^{(k)} = \sum_j W_{ij} f_j^{(k)} \tag{1.18}$$

where $f^{(k)}$ is the image estimate at iteration k.

MLEM Update Rule Starting from an initial estimate $f^{(0)} \geq 0$, the MLEM update is

$$f_j^{(k+1)} = f_j^{(k)} \frac{\sum_i W_{ij} \frac{p_i}{\hat{p}_i^{(k)}}}{\sum_i W_{ij}} \tag{1.19}$$

Equivalently, in matrix form:

$$f^{(k+1)} = f^{(k)} \odot \frac{W^\top \left(\frac{p}{W f^{(k)}} \right)}{W^\top \mathbf{1}} \tag{1.20}$$

where $\odot$ denotes elementwise multiplication and $\mathbf{1}$ is a vector of ones.

Algorithm steps

1. Initialize $f^{(0)}$ with nonnegative values.
2. Compute forward projection
$$\hat{p}^{(k)} = W f^{(k)} \tag{1.21}$$
3. Compute the ratio between measured and predicted data:
$$r_i^{(k)} = \frac{p_i}{\hat{p}_i^{(k)}} \tag{1.22}$$
4. Backproject the ratio:
$$b_j^{(k)} = \sum_i W_{ij} r_i^{(k)} \tag{1.23}$$

5. Normalize by sensitivity:

$$s_j = \sum_i W_{ij} \tag{1.24}$$

6. Update:

$$f_j^{(k+1)} = f_j^{(k)} \frac{b_j^{(k)}}{s_j} \tag{1.25}$$

Properties The MLEM shows different excellent features that spread the algorithm over multiple applications. First, it is derived from Poisson maximum-likelihood estimation. It guarantees nonnegativity of the reconstructed image. It produces high-quality reconstructions but may converge slowly. It is very sensitive to noise, indeed it is often combined with regularization (e.g., MAP-EM).

1.3.4 Ordered Subsets Expectation Maximization

The Ordered Subsets Expectation Maximization (OSEM) algorithm is an iterative method widely used for the reconstruction of tomographic images, particularly in modalities like PET and SPECT [32]. It is an accelerated version of the traditional Expectation Maximization (EM) algorithm, designed to improve convergence speed without sacrificing image quality. OSEM works by dividing the projection data into smaller subsets and updating the image estimate after processing each subset, rather than waiting for the entire dataset. This allows for faster convergence and reduced computational time, making it highly suitable for clinical applications where efficiency is critical [33]. The algorithm iteratively refines the estimated image to maximize the likelihood that the projected image matches the acquired data, incorporating the system's physical response and noise characteristics for improved accuracy [34]. For the effective implemented formula in the framework used for this study, the OSEM, that accelerates convergence by splitting the projection set into subsets S_j, has a rule derived at each iteration:

$$x^{(k+1)} = x^{(k)} \frac{A_{S_j}^* \left(\frac{y_{S_j}}{A_{S_j} x^{(k)} + b} \right)}{A_{S_j}^* \mathbf{1}} \tag{1.26}$$

Fig. 1.4 resume the concept of the OSEM: the sinogram is divide into subset, because it accelerates the convergence. So every subset follows the loop and the results are estimated at a number n of iterations that will be smaller then the MLEM number of iterations. The following Alg. 2 describes step-by-step the OSEM:

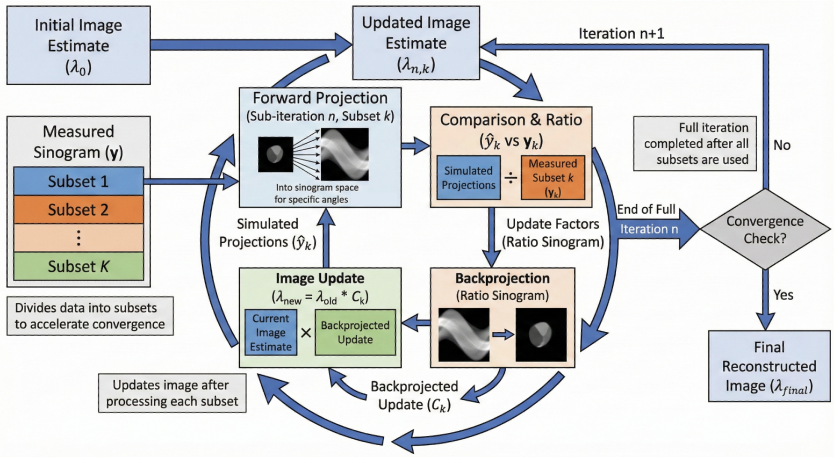


Figure 1.4: Explaining image for OSEM algorithm: here the λ correspond to x in the algorithm, C_k corresponds to ratio $c^{(s)}$. The figure resambles the MLEM algorithm with the change that the measured sinogram are divide in subsets to accelerate the loop.

Algorithm 2 Ordered Subsets Expectation Maximization (OSEM)

Require: Measured projections y, system matrix A, number of subsets S, initial image estimate $x^{(0)}$, number of iterations K

Ensure: Reconstructed image $x^{(K)}$

Partition the projection data into S ordered subsets: $\{y^{(1)}, y^{(2)}, \ldots, y^{(S)}\}$

for $k = 0$ to $K - 1$ **do**

  $x^{(k,0)} \leftarrow x^{(k)}$ ▷ Initialize iteration with previous estimate

  for $s = 1$ to S **do**

    Compute forward projection on subset s:

$$\hat{y}^{(s)} = A^{(s)} x^{(k,s-1)}$$

    Compute correction factor:

$$c^{(s)} = \frac{y^{(s)}}{\hat{y}^{(s)}}$$

    Backproject the scaled residual:

$$b^{(s)} = (A^{(s)})^\top c^{(s)}$$

    Normalize by system sensitivity:

$$d^{(s)} = (A^{(s)})^\top \mathbf{1}$$

    Update estimate for subset s:

$$x^{(k,s)} = x^{(k,s-1)} \cdot \frac{b^{(s)}}{d^{(s)}}$$

  end for

  $x^{(k+1)} \leftarrow x^{(k,S)}$ ▷ Store result for next iteration

end for

return $x^{(K)}$

1.3.4.1 MAP-Based OSEM and Resolution Recovery in Emission Tomography

In iterative image reconstruction for emission tomography, the Maximum A Posteriori (MAP) framework provides a principled way to incorporate both the statistical nature of the measured data and prior knowledge about the reconstructed image. When combined with Ordered Subsets Expectation Maximization (OSEM), MAP estimation enables faster convergence while controlling noise amplification through regularization, which is essential for achieving clinically useful image quality.

The reconstruction problem is commonly formulated by modeling the measured projection data $\mathbf{y}$ as a realization of a Poisson process, whose mean is given by $\mathbf{Ax} + \mathbf{r}$, where $\mathbf{x}$ denotes the unknown activity distribution, $\mathbf{A}$ is the system matrix encoding the physics of detection, and $\mathbf{r}$ accounts for additive background events such as scatter and randoms. In the MAP framework, the goal is to maximize the posterior probability $p(\mathbf{x}|\mathbf{y})$, which can be expressed, up to an additive constant, as the sum of the log-likelihood and the log-prior. This leads to the optimization problem

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}\geq 0} \{\log p(\mathbf{y}|\mathbf{x}) - \beta R(\mathbf{x})\} \tag{1.27}$$

where $R(\mathbf{x})$ is a penalty function encoding prior assumptions about the image, and β controls the strength of regularization. The penalty term is often chosen to encourage smoothness while preserving edges, as in quadratic, Huber, or total variation-based penalties.

OSEM accelerates the Expectation Maximization (EM) algorithm by partitioning the projection data into subsets and performing updates sequentially using each subset. In the MAP-OSEM setting, the standard EM update is modified by incorporating the gradient of the penalty term, resulting in an update equation that balances data fidelity and prior constraints. While OSEM improves convergence speed, it also introduces limit-cycle behavior and increased noise, making the role of the penalty term particularly important for stabilizing the reconstruction and ensuring convergence toward a meaningful solution.

An essential component of modern MAP-OSEM reconstruction is the explicit modeling and correction of the system's Point Spread Function (PSF). The PSF describes how a point source of activity is blurred by physical effects such as detector response, positron range, non-collinearity of annihilation photons, and depth-of-interaction effects. If uncorrected, these effects lead to spatially varying resolution loss and reduced contrast,

especially for small lesions. PSF correction is typically achieved by incorporating a spatially variant blurring model directly into the system matrix $\mathbf{A}$, often implemented as a convolution operation in image space or as a factorized component of the forward and backward projections.

By embedding the PSF model into the reconstruction process, MAP-OSEM effectively performs resolution recovery, allowing the algorithm to compensate for system blurring while remaining consistent with the measured data. This approach improves spatial resolution and contrast recovery without the need for post-reconstruction deconvolution, which is often unstable in the presence of noise. However, resolution recovery tends to amplify high-frequency noise, particularly near sharp edges, making penalized likelihood reconstruction crucial for controlling noise propagation.

Resolution-recovery algorithms based on PSF modeling must therefore be carefully balanced with appropriate regularization. The penalty term acts as a constraint that suppresses noise while allowing true structural details to emerge. In practice, the interaction between the PSF model, the number of OSEM subsets, and the regularization strength determines the final image characteristics. Overly aggressive resolution recovery can lead to edge artifacts and Gibbs-like ringing, whereas insufficient modeling of the PSF results in overly smooth images with poor lesion detectability. In summary, MAP-based OSEM reconstruction with penalized likelihood provides a flexible and powerful framework for modern emission tomography. The inclusion of PSF correction within the system model enables true resolution recovery, improving spatial resolution and quantitative accuracy. At the same time, the use of well-designed penalty functions is essential to mitigate noise amplification and ensure robust, clinically reliable images. Together, these elements form the foundation of state-of-the-art reconstruction algorithms used in PET and SPECT imaging.

Chapter 2

Deep learning methods: basic concepts

Deep learning is a subset of machine learning that uses artificial neural networks with multiple layers to learn and represent complex patterns in data. These deep neural networks are capable of automatically learning hierarchical features from raw input, making them particularly effective for tasks involving high-dimensional data such as images, speech, and text. The basics concepts of the deep learning are [35]:

1. *The hierarchical feature learning*: Each layer in the network learns increasingly abstract representations of the input data.
2. *End-to-end learning*: Deep learning models can learn directly from raw input to final output, often eliminating the need for manual feature engineering.
3. *Scalability*: Performance typically improves with more data and larger models, making deep learning well-suited for big data applications.
4. *Transfer learning*: Knowledge gained from one task can often be applied to related tasks, improving efficiency and performance.

Among the various deep learning architectures, Convolutional Neural Networks (CNNs) have emerged as a powerful tool, particularly in the domain of image analysis and computer vision.

2.1 Convolutional Neural Networks

CNNs are designed to automatically and adaptively learn spatial hierarchies of features from input images. The key components of a CNN include:

1. Convolutional layers: These layers apply learnable filters to the input, creating feature maps that highlight important patterns.
2. Pooling layers: These reduce the spatial dimensions of the feature maps, providing a form of translational invariance.
3. Fully connected layers: These layers typically appear at the end of the network and perform high-level reasoning.
4. Activation functions: Non-linear functions like Rectified Linear Unit (ReLU) that introduce non-linearity into the network, allowing it to learn complex patterns.

CNNs have several advantages that make them particularly suited for image analysis tasks, also in applications where one have to build the embedding space for a classification problem [36]. The main advantage is that Convolutional filters are applied across the entire image, significantly reducing the number of parameters compared to fully connected networks.
CNNs can capture local patterns and spatial relationships in images. And then, deeper layers in the network can learn increasingly complex and abstract features. And finally, the use of pooling layers helps the network recognize patterns regardless of their position in the image.
In medical imaging, CNNs have revolutionized various tasks including image classification, segmentation, and object detection. They have shown remarkable performance in analyzing complex medical images such as CT scans, MRIs, and X-rays, often achieving accuracy comparable to or exceeding that of human experts [37]. The milestone of the CNNs could be represented in Fig. 2.1, explaining how the convolution operation works.

2.1.1 Segmentation

In the image classification problem, the entire image is assigned to a class. A more detailed analysis is obtained when every pixel is assigned to one of a predefined set of classes, which means that the predicted space has the same dimensionality as the input space.

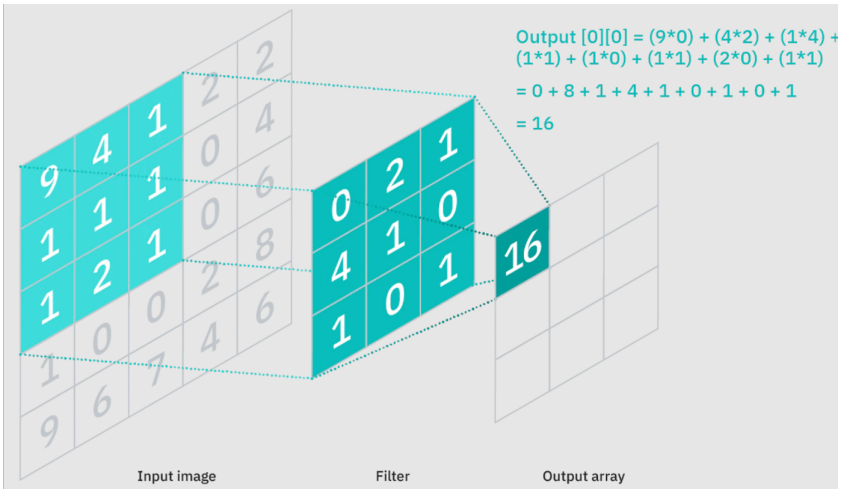


Figure 2.1: Convolution operation: terms coming from the operation of multiplying every function values by a kernel function. The kernel keep sliding over all the function values. At each position, the output is the value of every pixel of a new matrix. The output matrix is always smaller than the starting one. An image from course "Machine Learning".

The output will have C channels, which correspond to the class number, representing the probability for each class [38–40]. The predicted output of a segmentation network could be represented as a mask in which each pixel is coloured according to the class with the highest probability. In order to overcome the computational costs for applying convolutional filter to every patch centered on every pixel, one can uses a more efficient architecture by taking a deep convolutional network and adding layers that take the low-dimensional representation and develop it back up to the original image resolution. This operation is called Up-sampling. For the segmentation tasks this is shown to be best architecture. An example of segmentation for the CT images is in Fig. 2.2.

2.1.2 Object detection

In many applications of Computer Vision, deep learning could be used to determine the location of a determined object. The object is localized by using a bounding box like in Fig. 2.3, which is identified by the center's coordinates, the width and the height. The architecture extracts features through convolution and pooling layers, with fully connected layers followed by a SoftMax layer for classification [41, 42]. For localization, the final fully connected layer outputs bounding box coordinates. Detection models fall into two categories:

- Two-stage detectors, which generate region proposals before classification, including

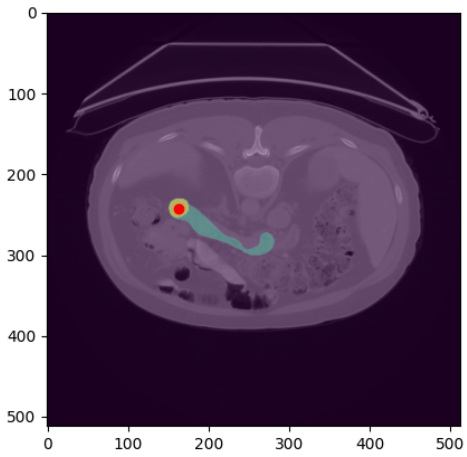


Figure 2.2: Example of segmentation of pancreas and tumour for a CT image. This is a validation image from segmentation task.

R-CNN, Fast R-CNN, Faster R-CNN, and Mask R-CNN, where R stands for Region-based. Indeed the model approach is to use a selective search algorithm to generate around 2,000 region proposals from an image. These proposals are likely to contain objects and are individually processed to detect and localize them more efficiently [43].

- One-stage detectors, treating detection as direct regression or classification tasks, like YOLO and SSD (Single-stage detectors) [44, 45]. In YOLO, that stands for You Only Look Once, the model looks only once in the network i.e. only one forward pass is required through the network to make the final predictions, because the model treats the detection as a single-regression problem, predicting bounding-boxes and class probabilities in one forward passage.

2.1.3 Reconstruction

The performance of deep learning algorithms relies on the quality of data used for model training. Higher-quality training data will become available with photon-counting CT scanners. At the same time, spectral data would greatly benefit from the computational abilities of deep learning methods. They achieves good performance by lowering image noise and artifacts while preserving spatial resolution and shortening reconstruction times, but its performance is only as good as the quality of the data used for its development

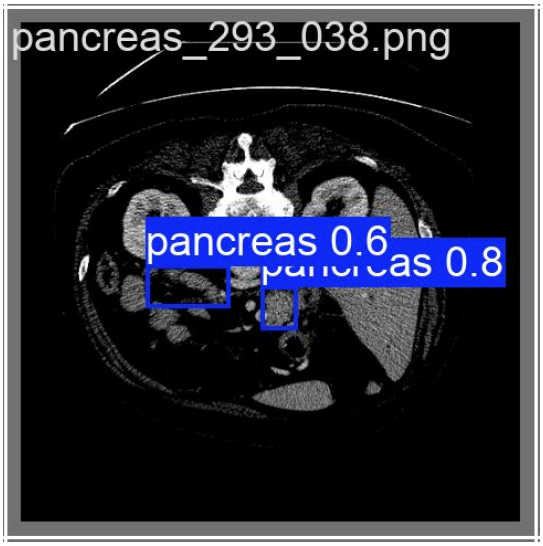


Figure 2.3: Example of object detection of pancreas in a CT image, with bounding boxes surrounding the pancreas. This is a validation image from localization task.

[46, 47].
The deep learning methods built for the reconstruction are divided by *direct* and *indirect*. In indirect either the filtered backprojection or iterative reconstruction is used. The three types of indirect frameworks are differentiated based on when the deep learning network is deployed. These are sinogram-based, image-based, and hybrid frameworks.

- Sinogram-based frameworks focus on sinogram optimization, and the network is deployed before the sinogram undergoes filtered backprojection or iterative reconstruction.
- In image-based frameworks, the network optimizes the image after initial reconstruction with filtered backprojection or iterative reconstruction.
- In hybrid frameworks, it is combined sinogram and image optimization.

In indirect reconstruction, deep learning networks are commonly CNN-based. Two examples are wavelet transform–based U-Net and residual encoder-decoder CNN. Transformer-based neural networks and generative adversarial networks are examples of a non-CNN–based approach to image optimization. This approach are commonly used for denoising problems, like in Fig. 2.4
In direct deep learning algorithms, it is reconstructed the sinogram directly into an image, without the use of filtered backprojection or iterative reconstruction. This potentially

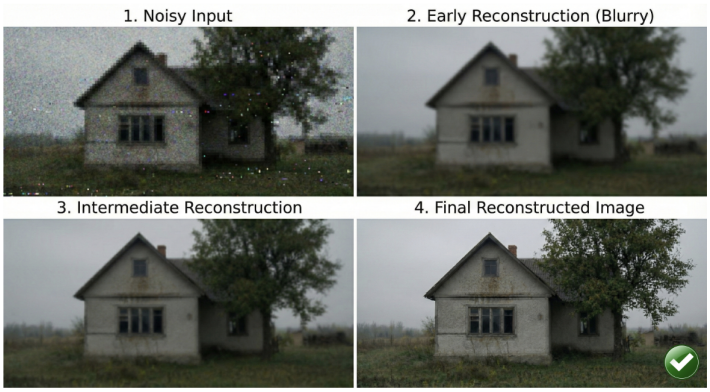


Figure 2.4: Example of denoising.

reduces artifacts introduced by classical approaches. However, this is only possible if the ground truth images used for model training do not contain classical methods-related artifacts. Examples of direct algorithms include Automated Transform by Manifold Approximation (AUTOMAP), which recasts image reconstruction as a data-driven supervised learning task that allows a mapping between the sensor and the image domain to emerge from an appropriate corpus of training data and iRadonMAP, that is an appropriative neural network, of which the architecture can be divided into two segments. In the first segment, a learnable fully-connected filtering layer is used to filter the radon projections along the view-angle direction, which is followed by a learnable sinusoidal back-projection layer to transfer the filtered radon projections into an image. The second segment is a common neural network architecture to further improve the reconstruction performance in the image domain [48, 49].

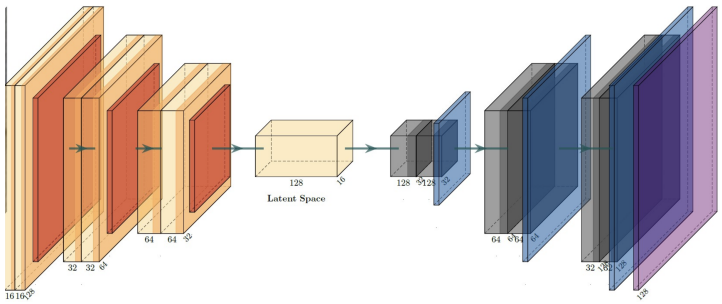


Figure 2.5: Example of AutoEncoder model for image generation [1].

[1] https://arc.fiu.edu/projects/artificial-intelligence-for-em-problem-set-dd/attachment/convolutional-autoencoder-cae-architecture/

2.1.4 Standard Autoencoder Architecture

The standard Autoencoder (AE) is an unsupervised learning model designed for efficient coding of input data, Fig. 2.5 [50, 51]. It is composed of two primary, interconnected networks:

- **Encoder (E):** This network takes the input data, $\mathbf{x}$, and progressively compresses it into a lower-dimensional representation, $\mathbf{z}$, known as the latent space (or bottleneck). The transformation can be represented as $\mathbf{z} = \mathbf{E}(\mathbf{x})$.
- **Decoder (D):** This network takes the latent representation, $\mathbf{z}$, and attempts to reconstruct the original input, $\hat{\mathbf{x}}$, such that $\hat{\mathbf{x}} = \mathbf{D}(\mathbf{z})$.

The objective function (loss function, $\mathcal{L}$) typically minimizes the reconstruction error:

$$\mathcal{L}_{AE} = \|\mathbf{x} - \hat{\mathbf{x}}\|^2 \tag{2.1}$$

The layers within both the encoder and decoder are typically structured as repeating blocks. A common configuration involves an alternation of:

1. **Convolutional Layers:** These perform feature extraction. For 3D volumetric data, this involves 3D Convolution operations. The encoder uses convolution with strides (or pooling) to reduce the spatial dimensions, while the decoder uses transposed convolutions (or upsampling) to increase them.
2. **Batch Normalization (BN) Layers:** Applied after the convolution, BN layers normalize the activations of the preceding layer. This stabilizes and accelerates the training process by reducing internal covariate shift.
3. **Activation Function:** A non-linear activation (e.g., ReLU or Leaky ReLU, which is the ReLU rectifier with a small positive slope in the negative side) is applied after the BN layer. A Leaky Rectified Linear Unit (Leaky ReLU) is defined as:

$$\text{Leaky ReLU}(\mathbf{x}) = \begin{cases} \mathbf{x} & \text{if } \mathbf{x} > 0 \\ a\mathbf{x} & \text{if } \mathbf{x} \leq 0 \end{cases} \tag{2.2}$$

where a is a small, fixed positive slope.

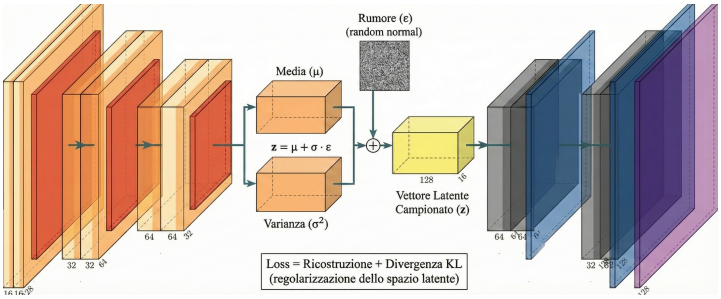


Figure 2.6: Example of Variational AutoEncoder model for image generation [52].

2.1.5 Variational Autoencoder

The Variational Autoencoder (VAE) is an extension of the standard AE with a key conceptual difference: the latent space is constrained to follow a specific probability distribution, typically a multivariate Gaussian distribution [53–55]. Fig. 2.6 is an example.

Instead of the encoder mapping the input $\mathbf{x}$ to a fixed point $\mathbf{z}$ in the latent space, it maps it to the parameters (mean $\boldsymbol{\mu}$ and standard deviation $\boldsymbol{\sigma}$) of a Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ in the latent space.

The latent vector $\mathbf{z}$ is sampled from this distribution using the reparameterization trick, in which I is the identity matrix:

$$\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}, \quad \text{where } \boldsymbol{\epsilon} \sim \mathcal{N}(0, I) \tag{2.3}$$

This modification makes the latent space continuous and allows for generative capabilities. The total loss function of the VAE is composed of the standard reconstruction loss and a regularization term (Kullback-Leibler divergence, D_{KL}) that forces the encoded distribution to be close to the prior ($\mathcal{N}(0, I)$):

$$\mathcal{L}_{VAE} = \mathcal{L}_{\text{Reconstruction}} + \beta \cdot D_{KL}(\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2) \parallel \mathcal{N}(0, I)) \tag{2.4}$$

2.2 U-Net and Transformer

The U-Net architecture is a type of convolutional neural network (CNN) introduced by Olaf Ronneberger [39]. It was designed specifically for semantic segmentation tasks in biomedical images, where the goal is to assign a class label to each pixel of an image (e.g., healthy vs. pathological tissue). What sets U-Net apart is its ability to deliver high-accuracy segmentation with a relatively small amount of labeled data, a common

limitation in medical domains [56]. The U-Net architecture is a powerful convolutional network originally developed for biomedical image segmentation. It is shown an example in Fig. 2.7. It is characterized by its distinctive symmetric, U-shaped design, which effectively combines high-level semantic information with low-level spatial detail [57–59]. The U-shaped architecture is composed of two main paths: a contracting (encoder) path and an expanding (decoder) path, connected by skip connections. The first part of the architecture is the *Contracting Path (Encoder)*. This part consists of convolutional layers followed by ReLU activations and max-pooling. This is the left side of the 'U'. It follows a standard convolutional network design where consecutive blocks of convolutional layers, followed by Batch Normalization, are interspersed with pooling operations (e.g., max-pooling or strided convolutions) that halve the spatial dimensions at every step while doubling the number of feature channel. The result is to extract hierarchical features and reduce spatial dimensions while increasing the number of feature channels. Going deeper, the receptive field increases, allowing the model to capture more global context. After the encoder, the architecture consists of an *Expanding Path (Decoder)*. This part performs up-convolutions (also called transposed convolutions or upsampling) to gradually restore the spatial resolution. This is the right side of the 'U'. It consists of upsampling layers (e.g., transposed convolutions) that double the spatial dimensions. At each level, the feature maps from the encoder are concatenated with the decoder features via skip connections. This structure allows the network to combine contextual information (from deep layers) with high-resolution localization (from shallow layers).
The final output layer is a 1×1 convolution, is applied at the end to map the multi-channel feature map to the desired number of classes, producing a segmentation mask where each pixel belongs to a specific class. The specific part of the network are the Skip Connections. The critical feature of the U-Net is the concatenation of feature maps from the contracting path to the corresponding layers in the expansive path. These skip connections transfer fine-grained spatial information lost during the pooling steps, enabling precise localization for segmentation tasks.

2.2.1 U-Net with Haar Wavelet Filters

A specialized variation, the U-Net with convolutional layers box implemented via Haar wavelet filters, leverages concepts from multiresolution analysis [62]. In Fig. 2.8 is reported this kind of architecture. The architecture consists of a layer that performs the Haar Wavelet Transform (HWT). The HWT decomposes a signal (or image) into different frequency sub-bands, which goes from the approximations to the details of the image. The

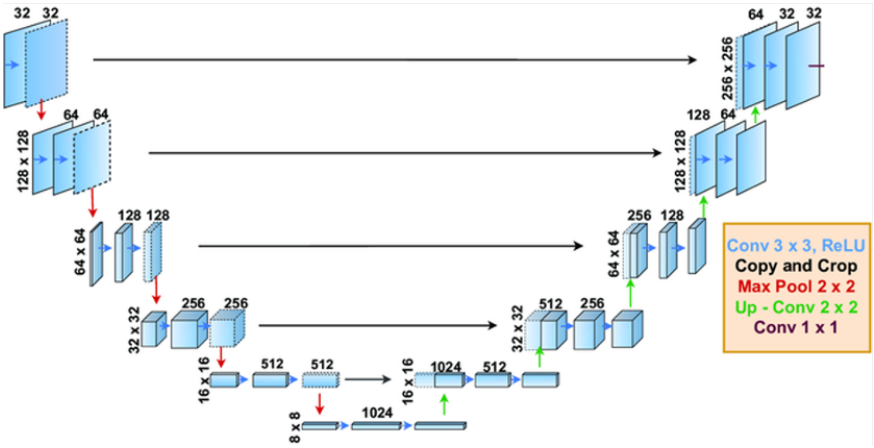


Figure 2.7: Example of a U-Net architecture used for medical image segmentation of Prostate cancer. The black arrows corresponds to the skip connections [60].

architecture can performs a Multiresolution Feature Decomposition. In this U-Net variant, the standard downsampling operation (max-pooling or strided convolution) is replaced or augmented by a structured decomposition using HWT. This allows the network to explicitly and mathematically capture features at different scales (multiresolution). The final path is the Reconstruction. The expansive path uses the inverse HWT or filters learned to perform multiresolution feature reconstruction. This approach can potentially offer improved performance in capturing both large-scale structures and fine-scale textures compared to standard pooling/upsampling, especially in tasks where spectral information is crucial.

2.2.2 Attention is all you need

The Transformer model, introduced in *"Attention Is All You Need"* in 2017, is built entirely around the Attention Mechanism, enabling highly parallelizable computation and an effective modeling of long-range dependencies within sequences. Unlike recurrent or convolutional architectures, the Transformer processes all tokens simultaneously and constructs contextualized representations by explicitly modeling pairwise interactions between tokens. The core component enabling this behavior is the Scaled Dot-Product Self-Attention mechanism, which assigns different levels of importance to tokens based on their relevance to one another [63].

Query, Key, and Value (QKV) Vectors For each input embedding $\mathbf{X}$, the Transformer learns three distinct linear projections that serve different roles within the attention

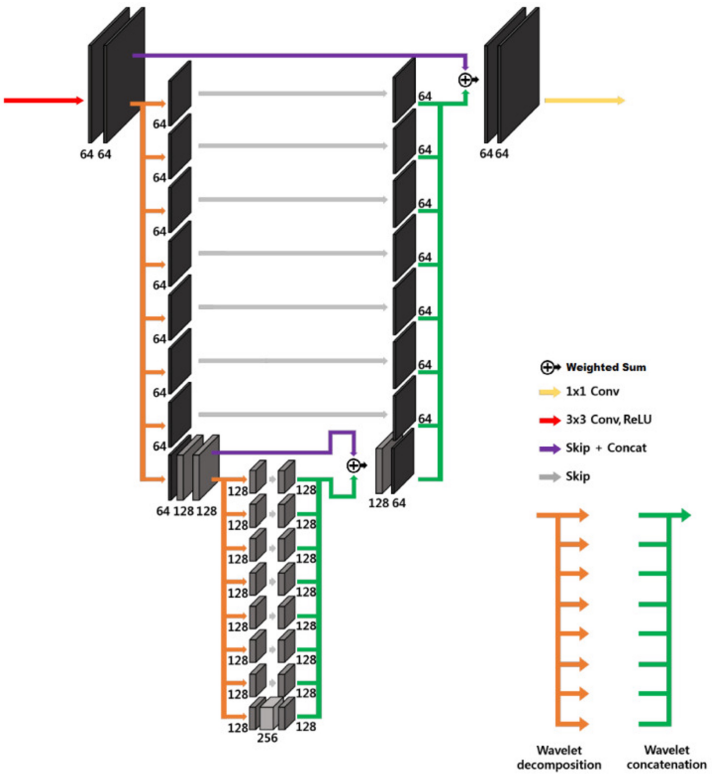


Figure 2.8: Example of U-Net tight frame architecture for an image of size = 64 × 64 × 64 [61].

mechanism. The query vector $\mathbf{Q}$ represents the information that a given token seeks from other tokens in the sequence, while the key vector $\mathbf{K}$ represents the information that each token can provide. The value vector $\mathbf{V}$ contains the actual content that is aggregated to form the output representation. These vectors are produced using learned projection matrices $\mathbf{W}^Q$, $\mathbf{W}^K$, and $\mathbf{W}^V$, respectively:

$$\mathbf{Q} = \mathbf{X}\mathbf{W}^Q \quad \mathbf{K} = \mathbf{X}\mathbf{W}^K \quad \mathbf{V} = \mathbf{X}\mathbf{W}^V \tag{2.5}$$

Scaled Dot-Product Attention Formula The computation of attention scores begins by measuring the similarity between query and key vectors through a dot product operation. These raw similarity scores are then scaled by the inverse square root of the key dimensionality $\sqrt{d_k}$ to prevent excessively large values that could negatively affect gradient stability during training. A Softmax function is subsequently applied to normalize the scores into attention weights, which are finally used to compute a weighted sum of the value vectors. This entire process is compactly expressed by the

Scaled Dot-Product Attention formulation [64, 65]:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \tag{2.6}$$

2.2.3 Multi-Head Attention

Multi-Head Attention extends the basic attention mechanism by executing multiple attention operations in parallel, allowing the model to attend to information from different representation subspaces simultaneously. Each of the h attention heads applies separate learned projections to the input, enabling the capture of diverse relationships such as syntactic, semantic, or positional dependencies [66]. The outputs of all attention heads are concatenated and passed through a final linear transformation to generate the output representation, as illustrated in Fig. 2.9:

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)\mathbf{W}^O \tag{2.7}$$

where each individual head is computed as:

$$\text{head}_i = \text{Attention}(\mathbf{Q}\mathbf{W}_i^Q, \mathbf{K}\mathbf{W}_i^K, \mathbf{V}\mathbf{W}_i^V) \tag{2.8}$$

Transformer Architecture Components The original Transformer architecture is composed of stacked Encoder and Decoder layers, both of which employ residual connections and Layer Normalization to facilitate stable and efficient training.

Encoder Layer Each Encoder layer consists of a Multi-Head Self-Attention module followed by a position-wise Feed-Forward Network. The feed-forward network is a two-layer fully connected structure that is applied independently and identically to each token representation, enabling nonlinear feature transformation at the token level.

Decoder Layer Each Decoder layer is composed of three sub-layers. The first is a masked Multi-Head Self-Attention mechanism that prevents each position from attending to future tokens, thereby preserving the autoregressive property during training. This is followed by an Encoder-Decoder, or cross-attention, mechanism in which the queries originate from the decoder while the keys and values are obtained from the final encoder outputs, allowing the decoder to selectively focus on relevant portions of the input sequence. The final sub-layer is a position-wise Feed-Forward Network, analogous to that used in the encoder.

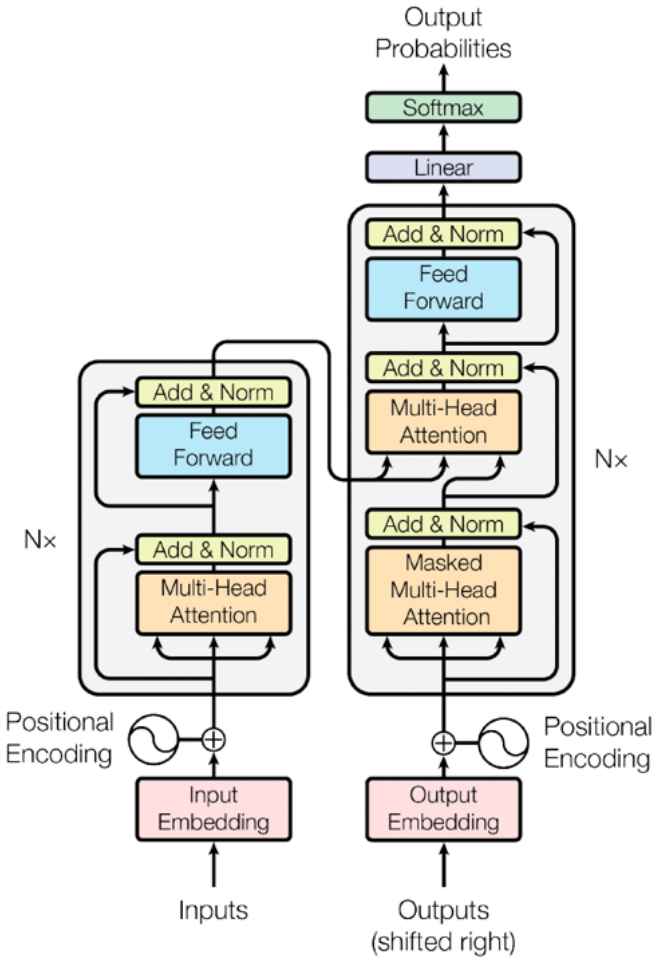


Figure 2.9: Example of transformer [63].

Positional Encoding Because the attention mechanism itself is permutation-invariant and does not encode sequence order, positional information is injected into the model through Positional Encoding. In the original Transformer, this is achieved using fixed sinusoidal functions that are added to the input embeddings, allowing the model to infer relative and absolute token positions [67].

2.2.4 Applications in Computer Vision

Transformers have been successfully adapted to computer vision tasks by reformulating images as sequences of tokens. The Vision Transformer (ViT) is a prominent example, in which images are decomposed into patches and processed using standard Transformer encoders [68].

Vision Transformer (ViT) Paradigm In the ViT framework, an input image is first divided into a set of non-overlapping patches, each of which is flattened and linearly projected into a fixed-dimensional embedding space. These patch embeddings are treated as tokens and augmented with positional encodings to preserve spatial information. The resulting sequence is then processed by a Transformer encoder, which uses self-attention to model global contextual relationships across the entire image.

Image Detection The Detection Transformer (DETR) reformulates object detection as a direct set prediction problem, removing the need for handcrafted components such as anchor boxes and non-maximum suppression [69, 70]. In this approach, a convolutional neural network backbone extracts feature maps that are passed to a Transformer encoder. The decoder receives a fixed set of learned object queries and employs cross-attention to focus on relevant image features. The model produces a fixed number of predictions in parallel, each corresponding to an object class and bounding box, and is trained using a set-based loss function that relies on bipartite matching, such as the Hungarian algorithm, to align predictions with ground-truth annotations.

Image Segmentation Transformers have also demonstrated strong performance in image segmentation tasks by leveraging their ability to capture global context [71–73]. In semantic segmentation, a Transformer encoder generates contextualized patch representations that are subsequently upsampled by a segmentation head to recover pixel-level predictions. For instance and panoptic segmentation, architectures such as MaskFormer introduce learnable mask queries within the decoder. Each query predicts both a semantic class and a mask embedding, which is combined with high-resolution feature maps to generate instance-specific binary masks, thereby unifying multiple segmentation tasks within a single framework.

Chapter 3

Development of machine vision algorithms for medical images

The medical imaging became a useful way to improve the medical therapy in case of cancer. Several methods based on imaging records were developed during the years, that are called "machine vision algorithms", that enables the computer to *see* as an human made for different applications. Machine vision algorithms have become a fundamental component of modern image analysis, enabling computers to automatically interpret and extract meaningful information from visual data. By taking advances in signal processing, pattern recognition, and machine learning, machine vision systems aim to replicate and, in some cases, surpass human visual perception for specific analytical tasks. These algorithms are widely applied across domains such as medical imaging, remote sensing, industrial inspection, and autonomous systems, where large volumes of image data must be analyzed in a consistent and reproducible manner.

A central objective of machine vision is the transformation of raw image data into quantitative descriptors that capture relevant structural, textural, and statistical properties of the observed objects or regions. Traditional approaches rely on handcrafted features derived from intensity distributions, edge information, shape descriptors, and texture measures, while more recent methods incorporate data-driven representations learned through deep neural networks.

The radiomics workflow commonly involves image acquisition, preprocessing, region-of-interest delineation, feature extraction, and subsequent analysis using statistical or machine learning models. Machine vision algorithms play a crucial role at multiple stages

of this pipeline, particularly in image normalization, automated segmentation, and robust feature computation.

These techniques are spread between the basic image processing and the advanced deep learning for segmentation of tissue. The key areas are object detection, image classification and segmentation and feature extraction. In this chapter we will focus on a combined task: localization and segmentation.

3.1 One step segmentation of pancreas

Pancreatic cancer is one of the leading causes of cancer related deaths worldwide. Pancreatic tumors represent a heterogeneous group of neoplasms arising from the exocrine or endocrine components of the pancreas. They are characterized by complex molecular alterations, aggressive biological behavior in several subtypes, and significant challenges in diagnosis and management. This text provides a scientific overview of pancreatic tumors with emphasis on pathology, molecular mechanisms, and clinical implications.

The pancreas is a dual-function organ composed of exocrine acinar and ductal cells as well as endocrine islet cells. Tumors of the pancreas originate from these distinct cellular compartments and therefore exhibit diverse histological and functional characteristics. Pancreatic tumors are of particular scientific interest due to their often late presentation, resistance to therapy, and high impact on patient outcomes.

Pancreatic tumors are broadly classified into exocrine and neuroendocrine neoplasms. Exocrine tumors account for the majority of cases and are most commonly represented by pancreatic ductal adenocarcinoma, which arises from the ductal epithelium. These tumors are typically characterized by dense stromal tissue, infiltrative growth patterns, and marked cellular atypia. In contrast, pancreatic neuroendocrine tumors originate from hormone-producing islet cells and display variable differentiation and proliferative activity, which influence their biological behavior.

At the molecular level, pancreatic tumors are associated with the accumulation of genetic and epigenetic alterations that disrupt normal cell cycle regulation, apoptosis, and tissue architecture. Frequently affected signaling pathways include those involved in growth factor signaling, DNA damage response, and cellular metabolism. The tumor microenvironment plays a critical role in disease progression, as interactions between

neoplastic cells, stromal fibroblasts, immune cells, and extracellular matrix components contribute to tumor growth and therapeutic resistance.

The clinical manifestations of pancreatic tumors are often nonspecific and may include metabolic or digestive disturbances depending on tumor location and functional status. From a diagnostic perspective, imaging modalities such as computed tomography and magnetic resonance imaging are used to assess tumor size, extent, and anatomical relationships. Histopathological examination, supported by immunohistochemical and molecular analyses, is essential for definitive diagnosis and classification.

Management strategies for pancreatic tumors depend on tumor type, stage, and molecular profile. Surgical resection remains a key approach for localized disease, while systemic therapies are employed in more advanced cases. Ongoing research focuses on understanding tumor biology at the molecular level to identify biomarkers and therapeutic targets that may improve early detection and treatment efficacy.

Since patients rarely exhibit symptoms until the disease reaches an advanced stage, early diagnosis is critical for improving the five-year survival rate. Furthermore, pancreatic masses encompass various types, presenting a management dilemma as some are precancerous while others carry minimal risk of developing invasive cancer.
Computed tomography (CT) is widely used for initial evaluation of cancer lesions. In CT images, tumor areas are meticulously identified and delineated by expert radiologists. In certain clinical scenarios, combining information from different imaging modalities, such as CT and digital radiography, can provide a more comprehensive view of the anatomy and pathology. However, there are at least three problems concerning the segmentation of pancreatic tumoral masses, which can reduce the quality of results. The first is related to tissue density: cystic ducts can be easily confused with pancreatic ones. The second issue is that pancreatic masses vary in size, shape, and texture. Additionally, since they occupy only a small portion of the abdominal CT image, physicians need to explore wide part of the slices to identify them. However, the main problem is due to the blurry and indistinct pancreatic borders. The third problem lies in the multitude of pancreatic tumor types, such as serous cystic neoplasm (SCN), mucinous cystic neoplasm (MCN), intraductal papillary mucinous neoplasm/carcinoma (IPMN/IPMC), ductal adenocarcinoma (PDAC), pancreatic pseudocyst (PPC), pancreatic neuroendocrine carcinoma/tumor (PNEC/PNET).
Amid the strides made in modern medicine, including the development of enhanced

surgical systems [74–76], diagnostic techniques [77] and sensing techniques [78], Deep Convolutional Neural Networks (DCNN) have emerged as an accurate and efficient approach for tumor segmentation in medical images. DCNNs excel at identifying indistinguishable pixels or subtle changes that may be overlooked by the human eye, enabling the construction of more complex feature structures [79, 80].
Pancreatic tumor segmentation is also being pursued using fully convolutional DCNNs [81]. In particular, studies have started focusing specifically on the U-Net architecture, which has brought a significant improvement in segmentation accuracy [82]. Zhou et al. have implemented a deep learning-based method that first segments the pancreas through a dedicated DCNN model, and then crop, resize, and manage the original image to segment the pancreatic cyst through another model [83]. Guo et al. have proposed a 3D tumor segmentation using a combination of a U-Net architecture and a graph-based network [84]. Initially, tumor tissue is manually retrieved from the total CT image and saved as a cube centered on the tumor. It is then sliced into 2D images, and every slice and its two adjacent slices are provided to a U-Net model to obtain segmentation, which is used as a probability map. Ultimately, this map serves as the input for a graph-based algorithm that performs the final segmentation.

In this work, a one-step segmentation framework is employed in which both the pancreas and the PDA tumor are segmented directly from CT images in a single unified process. Unlike multi-stage approaches that first localize the pancreas and then segment the tumor, the proposed one-step method reduces pipeline complexity in order to minimize the error propagation.

Prior to segmentation, several pre-processing operations are applied to the CT images to enhance image quality and improve model performance. These operations include windowing on the scale of intensity that is applied to improve the visibility of the pancreas and tumor regions, enabling the neural network to better distinguish relevant anatomical structures from surrounding tissues; intensity normalization to reduce variations across scans acquired from different imaging devices and protocols and spatial resampling to ensure consistent resolution across all images. In addition, some operations like rotations, translations and brightness are applied to augment the dataset.

The segmentation task is performed using an attention-based deep neural networks, typically built upon an encoder–decoder architecture such as the U-Net. The encoder extracts hierarchical feature representations from the input CT images, while the decoder

reconstructs dense pixel-wise segmentation maps. Attention mechanisms are integrated into the network to guide the model toward clinically relevant regions by selectively emphasizing features associated with the pancreas and PDA while suppressing irrelevant background information. By incorporating attention gates into the skip connections, the networks retain fine spatial details while reducing the influence of noisy or non-informative features.

The networks produce a pixel-level segmentation map that identifies the PDA region in a single forward pass. This one-step attention-based approach offers several advantages, including reduced computational complexity, improved segmentation accuracy for small and low-contrast tumors, and faster inference times.

3.1.1 Databases

In this thesis, only CT imaging data from public dataset are used. Research has been conducted to identify suitable data sources, i.e., datasets containing CT images annotated with semantic segmentation of pancreas and pancreatic tumors. The annotation is made by physicians, with help of segmentation software.
Two main datasets have been found: the Medical Segmentation Decathlon (MSD) [85] and the Clinical Proteomic Tumor Analysis Consortium Pancreatic Ductal Adenocarcinoma Collection (CPTAC-PDA) [86], which are described below. The MSD database comprises various organ images and is used as database for segmentation challenges. The data set was designed to explore the axis of difficulties typically encountered when dealing with medical images, such as small data sets, unbalanced labels, multi-site data and small objects. It offers different repositories, among which is the Pancreas database, namely Task 07, containing a total of 420 3D CT scans divided into the training and test parts. The training subset consists of 281 images with associated labels for tumors and pancreas, manually segmented by radiologists. In contrast, the test subset contains 139 CT images without labels. CT images in the MSD database exhibit varying numbers of slices on the transverse plane, each with dimensions of 512×512 pixels.

The MSD Task 07 dataset has been utilized for training as it contains more samples and provides annotations for both the pancreas and pancreatic tumor. Pancreas segmentation proves particularly useful during training, aiding the neural network in learning to recognize the pancreas's location and consequently enhancing its ability to identify pancreatic cancer lesions.

The CPTAC collects and distributes anonymized medical data from patients diagnosed with various types of cancer, including pancreatic ductal adenocarcinoma (PDA), with the goal of understanding cancer's molecular basis through proteome and genome analysis. These datasets comprise proteomic, genomic, imaging and clinical information. Radiology and pathology images are available in the Cancer Imaging Archive, facilitating research into cancer phenotypes correlated with proteomic, genomic, and clinical data.
The imaging cohort of CPTAC-PDA comprises 168 human subjects diagnosed with PDA. Radiology images are gathered from standard-of-care imaging performed on patients immediately before the pathological diagnosis, and from follow-up scans where available. Consequently, the radiology image data is heterogeneous in terms of scanner modalities, device manufacturers, and acquisition protocols. Specifically, radiology images (in DICOM format) are provided for only 108 of the patients, while pathology images are provided for all the patients. Recently, manual annotations for the radiology images have been also made available, unleashing the potential for utilizing CPTAC-PDA for neural network training and validation.

Radiology images have been meticulously filtered and selected to include only CT images of patients with PDA. Despite the collection encompassing magnetic resonance (MR), positron emission tomography (PET), and ultrasound (US) medical images, these were excluded to ensure data coherence, given that the MSD dataset exclusively comprises CT images. Moreover, some CT images pertain to secondary cancerous lesions in other body parts, which have been automatically identified and removed by inspecting the annotations in Digital Imaging and Communications in Medicine (DICOM) files defining regions of interest (ROIs) relative to cancer lesions. Following this filtering process, 134 CT images corresponding to 87 patients have been selected.

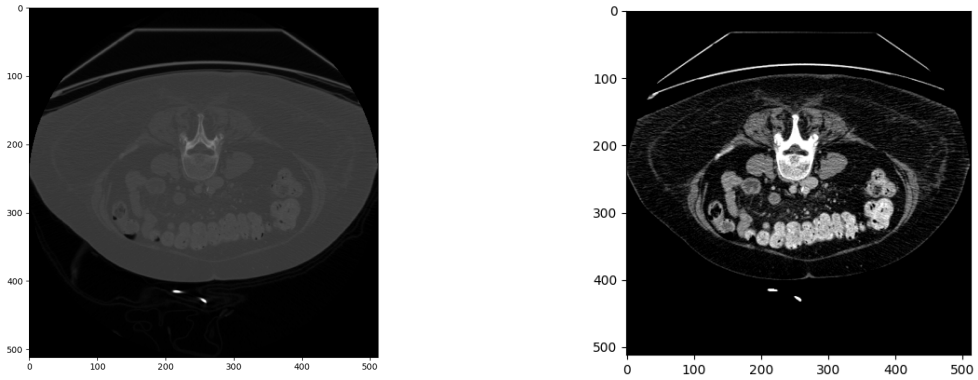
The CPTAC-PDA dataset is used as test dataset, to test the generalization ability of trained models on a completely independent dataset.

3.1.2 Preprocessing operations

Some preprocessing operations have been applied to both the MSD and CPTAC-PDA datasets. Initially, considering that CT images are acquired with instruments presenting varying spatial resolutions, they are resampled to a consistent spatial resolution of 1 mm in the xy plane (transverse plane). Slices in this plane are then fed to neural network models for 2D segmentation and require a uniform resolution. Resampling on the z-axis,

with the same spacing of 1 mm, has been applied for some of the trainings only, to assess its impact on the final segmentation accuracy of the model.

CT images are cropped on the xy plane to a dimension of 256×256 pixels, ensuring that the resulting images contain both the pancreas and pancreatic tumor in all cases. Subsequently, windowing is applied to enhance contrast by clipping intensity values in the range [-100, 240], typically employed in the literature [87], and visible in Fig. 3.1



(a) (b)

Figure 3.1: (a) The raw image in the dataset without windowing. (b) The result of the windowing process

.

where is shown the image before and after the windowing. Finally, images are normalized according to one of the following two considered approaches: (i) min-max normalization in the range [0, 1], (ii) z-score normalization.

Only slices on the transverse plane containing regions of pancreas and/or pancreatic cancer in the annotation are used for training and testing.

3.1.3 Data handling

In this study, the use of 2.5D models for CT image segmentation is investigated, as a upgraded approach to 2D model. This approach involves providing the neural network with adjacent slices to the one being analyzed. Consequently, the neural network can leverage 3D information without the complexity of a full 3D segmentation model [88, 89]. Specifically, this study examines the use of 11 input slices, comprising the central slice and 5 slices from each side.

Augmentation The augmentation of the dataset is not mandatory because the number of images is 420 CT scan and the model is quite small. In any case, the pipe-line was developed also with the augmentation data flow. The augmentation techniques used are three: the transformation of the image by rotations and translations, and the brightness over all the image. The rotation pick up a center between the size of the image, and apply the rotation of a random angle. The translation, took the center previously used for the rotation and make a traslation of a vector that is randomly choosen from an fixed interval. The brightness take the normalized image and add a constant value to the image [90].

Leakage In Machine Learning analysis is very common find leakage in data ingestion. This kind of problem is the use of information during the training phase that would not be available during prediction time (or testing time). This could happen in several ways: one is the *Group Leakage* [91], that occurs when in a database are not included a grouping split column. In practice, an example with images is described in the following: the dataset is made by 100k x-rays of 30k patients, meaning 3 images per patient. The dataset split mode is random splitting instead of ensuring that all images of a patient were in the same split [92]. Hence the model partially memorized the patients instead of learning to recognize pneumonia in chest x-rays, and the model appears to perform better during evaluation, but when the model is tested in a production environment the performance are totally lower than the evaluation performance [93, 94].
In this study the dataset is first divided by slice, than by patients. The results for the two modalities are reported. They evidence that the leakage issue is related to the presence of contiguos images from patients in the training set and in the test set, that force the model to memorize them.

3.1.4 Models and training

The MSD CT images are utilized for training, with 20 % of them reserved for validation/test purposes. The neural network models are developed using the Python language and, specifically, the TensorFlow framework. Training of each model is carried out for 50 epochs employing the Adam optimizer and using the cross-entropy as loss function. The ReCaS computing center in Bari, equipped with NVIDIA A100 GPUs, is used as training platform. Augmentation techniques are utilized in some of the training processes, involving random shift of the slices in the x and y directions, random rotations on the xy plane, and random brightness adjustments. To ensure full reproducibility of the experi-

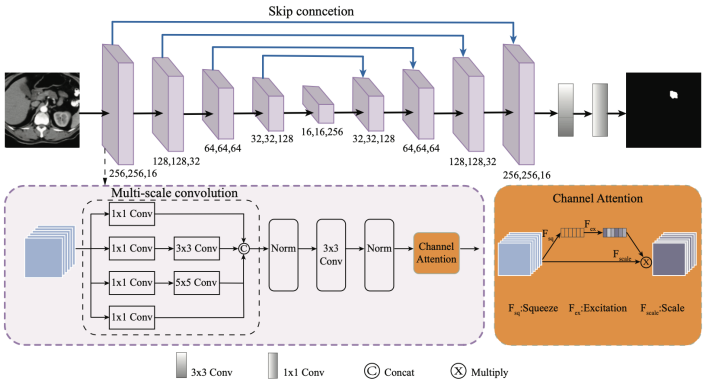


Figure 3.2: MSCA model [88].

ments, detailed training configurations are reported. The dataset was rigidly partitioned into training and validation sets. Specifically, patient-level splitting was strictly enforced to prevent data leakage, ensuring that slices from the same patient did not appear in multiple subsets. A 5-fold cross-validation was employed to robustly estimate model performance variability. The neural network models are developed using the Python language and, specifically, the TensorFlow framework. Training of each model is carried out for a maximum of 50 epochs using the Adam optimizer with an initial learning rate of 1×10^{-4} and a cross-entropy loss function. A learning rate decay schedule was applied, reducing the rate by a factor of 0.1 upon plateauing. To prevent overfitting, an early-stopping criterion was implemented, which halted the training if the validation loss did not improve for 10 consecutive epochs. The ReCaS computing center in Bari, equipped with NVIDIA A100 GPUs, is used as training platform. Augmentation techniques are utilized in some of the training processes, involving random shift of the slices in the x and y directions, random rotations on the xy plane, and random brightness adjustments.

MSCAnet The model is called Multi-Scale Channel Attention U-Net and is an embedding of the multi-scale network into the decoder, in order to provide supplemental information to overcome the loss of information in the upsampling and the drift of the localized tumor due to the upsamling and skip connections. The channel attention unit after the multi-scale convolution was adopted to emphasize the informative channels, which was observed to have the result to accelerate the positioning process. The described model is shown in Fig. 3.2.

SRSnet The DCNN model used in this study for 2D segmentation of CT medical images is based on the recently introduced SRSNet. It consists of a U-Net derived encoder/decoder

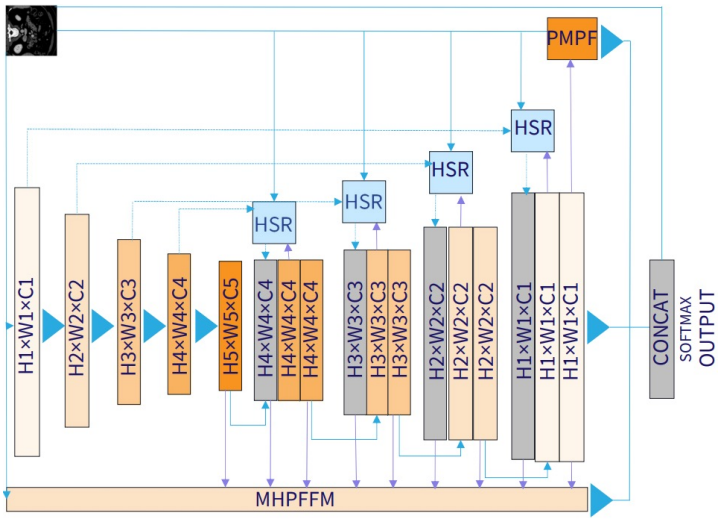


Figure 3.3: Schematic representation of the SRSNet architecture used for the 2D segmentation task. The model incorporates a Multi-scale High-resolution Pre-segmentation feature fusion mechanism (MHPF) and replaces standard skip connections with high-resolution spatial information recovery (HSR) modules to minimize information loss [95].

architecture, with high-resolution spatial information recovery mechanism (HSR) replacing the skip connections typically connecting encoder and decoder. The network output is derived from the input image and three pre-segmented features, combined by the use of the Multi-scale High-resolution Pre-segmentation feature fusion mechanism (MHPF), a pyramid multi-scale feature perception and fusion mechanism. In Fig. 3.3 the model is shown. Additionally, in contrast to the U-net architecture, which includes pooling layers associated with high feature information loss, SRSNet employs convolutional layers that consider the learning state. Furthermore, up-sampling layers are replaced by deconvolution layers. This approach enables the backbone network to learn features with minimum information loss, resulting in a more accurate pre-segmented map with high-resolution features preserved.

MDAGnet The Multi-dimensional Attention Gate Net (MDAG) is proposed to combine three attention mechanisms, spatial, channel and multi-dimensional feature map. The MDAG is added to the skip connections, which extracts the channel attention coefficients and spatial attention coefficients from the high-resolution feature map. After that, the double attention coefficients are multiplied by the shallow feature map to suppress the irrilevant regions, while are enphasized the salient features usueful for pancreas segmentation. In Fig. 3.4 is shown the model described.

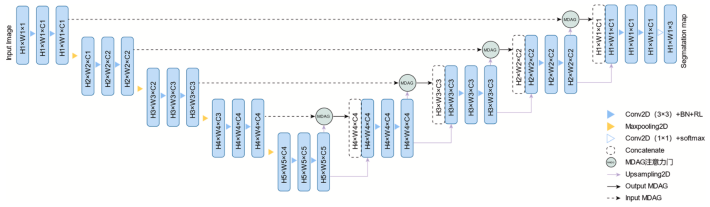


Figure 3.4: Architecture of the Multi-dimensional Attention Gate Net (MDAG). The model integrates spatial, channel, and multi-dimensional feature map attention mechanisms within the skip connections. This design extracts dual attention coefficients to suppress irrelevant background regions while emphasizing salient features crucial for accurate pancreas segmentation [96].

Attention-Res-U-Net The proposed architecture is an Attention-Res-U-Net, a specialized deep learning model developed for semantic segmentation tasks in medical image analysis. This architecture combines three key concepts, namely the U-Net encoder–decoder structure, residual learning, and attention mechanisms. The model follows the classical U-shaped design composed of a contracting path and an expansive path; however, unlike the standard U-Net, conventional convolutional blocks are replaced with residual blocks to enable more effective training of deeper networks. In addition, attention gates are incorporated into the skip connections to selectively filter spatial information passed from the encoder to the decoder.

The encoder, or contracting path, is responsible for extracting hierarchical feature representations from the input image. Instead of using stacked convolutional layers, the encoder employs residual convolutional blocks (`ResConvBlock`), each consisting of two 3×3 convolutional layers followed by batch normalization and a nonlinear activation function such as ReLU or Tanh. A residual shortcut connection adds the block input directly to its output, which helps mitigate the vanishing gradient problem and preserves low-level spatial information. Between successive residual blocks, spatial downsampling is performed using max pooling operations with a kernel size of 2×2, which halve the spatial resolution while simultaneously increasing the number of feature channels. As a result, the encoder progressively increases its representational capacity, starting from a base number of filters (e.g., 128) and reaching a bottleneck with up to 2048 feature channels.

A key limitation of the standard U-Net architecture is that skip connections indiscriminately transfer all encoder features to the decoder, including irrelevant background information. To address this issue, the proposed model integrates attention mechanisms

into the skip connections. Each attention block receives two inputs: the feature map from the encoder at the corresponding level and a gating signal from a deeper decoder layer that provides coarse, high-level contextual information. Both inputs are first projected into an intermediate feature space using 1×1 convolutions. The transformed features are then combined and passed through a nonlinear activation function, followed by a sigmoid function, to generate an attention map with values in the range $[0, 1]$. This attention map is applied to the encoder feature map through element-wise multiplication, effectively suppressing irrelevant regions while enhancing features that are most informative for the segmentation task.

The decoder, or expansive path, reconstructs the segmentation mask by progressively restoring spatial resolution. Feature maps are upsampled using bilinear interpolation by a factor of two and concatenated with the corresponding encoder features that have been filtered by the attention blocks. This gated concatenation ensures that only relevant spatial information contributes to the reconstruction process. The combined features are then refined using residual convolutional blocks, which help recover fine structural details and improve boundary delineation in the segmentation output.

Finally, the output of the last decoder block is passed through a 1×1 convolutional layer to map the extracted features to the desired number of output classes. For binary segmentation tasks, a sigmoid activation function is applied to generate a probability map representing the likelihood of each pixel belonging to the target class, such as a tumor region. A schematic representation of the Attention-Res-U-Net architecture used in this work is shown in Fig. 3.5.

3.1.5 Results and Discussion

The results obtained are evaluated with the most common metric in pancreas and PDA segmentation, the DICE score (DSC):

$$DSC = \frac{2|A \cap B|}{|A| + |B|} \tag{3.1}$$

where, A and B are examples of image tensors, and the $\cap$ operation means the multiplication of the two tensors. This metric is innerly linked to the number of pixels that are detected belonging to tumour or not.

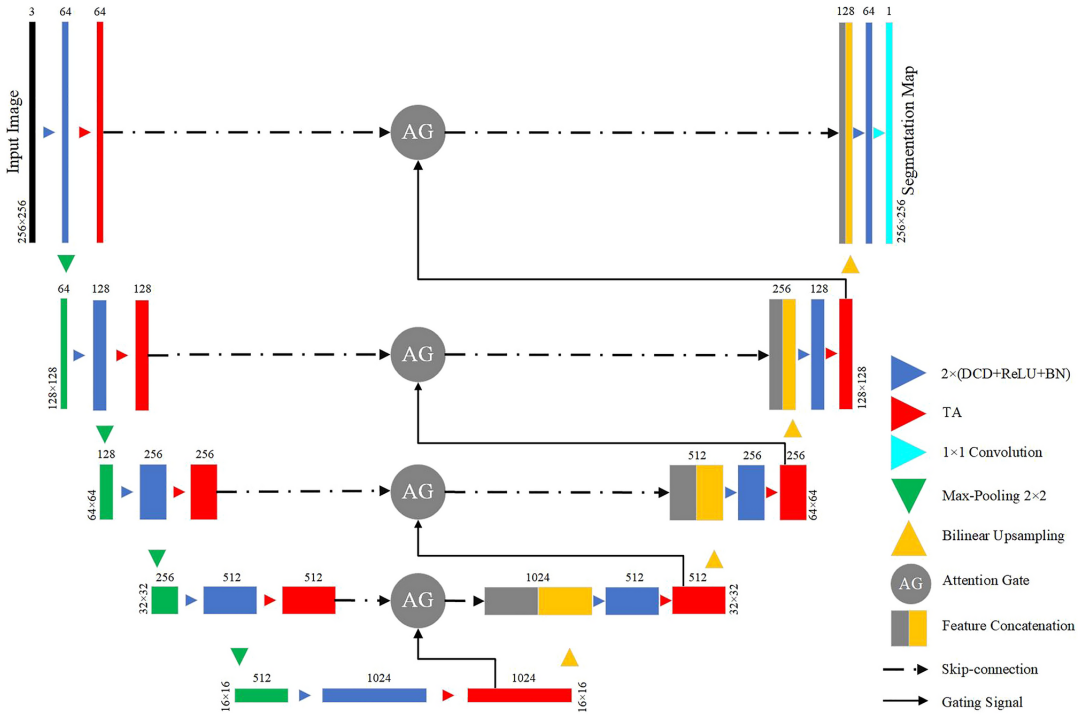


Figure 3.5: Attention-Res-U-Net model. The combination DCD (Deep Convolution Decoder), ReLU (Rectified Linear Unit) and BN (Batch Normalization), with Tensor Attention followed by Upsample, is reported in the model [97]. In the DCD is built also the ResConvBlock.

To assess the statistical robustness of the proposed pipelines, performance metrics are reported as Mean ± Standard Deviation across the cross-validation folds. A paired t-test was conducted to compare the 1-step and 2-step segmentation approaches. The improvement yielded by the 2-step pipeline was found to be statistically significant ($p < 0.05$), confirming the consistency of the localization-first strategy. Furthermore, 95% Confidence Intervals (CIs) were computed for the primary metric (DSC), highlighting the variability introduced by different random initializations and patient anatomical heterogeneities.

Results are reported in Tab. 3.1, where they are summarized according to the model employed and the corresponding preprocessing configuration. For each experiment, the set of segmentation labels used is also specified. In particular, the models were trained either using both pancreas and tumor labels simultaneously or using a single label per experiment.

For the MSCA_v03 configuration, as well as its corresponding variant with suffix

_1, a comparative analysis can be performed between pancreas-only segmentation and multi-class segmentation involving both the pancreas and the tumor. The results indicate that segmenting the pancreas as a single class yields superior performance compared to the multi-class approach. Furthermore, the inclusion of the data augmentation technique referred to as "shift" leads to improved segmentation performance, as evidenced by the comparison between MSCA_v04 and MSCA_v06. In contrast, the application of brightness augmentation results in a degradation of performance, with a noticeable reduction in the Dice coefficient observed for MSCA_v07.

In the MDAG study, a clear distinction can be observed between one-class and multi-class segmentation models. In particular, a substantial difference in tumor Dice scores is evident when comparing the MDAG_v02 and MDAG_v03 configurations, highlighting the impact of class formulation on tumor segmentation performance.
The comparative analysis of the segmentation models highlights how architectural choices, preprocessing strategies, and contextual information jointly influence the ability to delineate both pancreatic tissue and tumour regions. The SRSNet experiments reveal a strong dependence on the type of intensity normalization applied to the input CT volumes. Models trained with simple min–max scaling exhibit more stable behaviour than those relying on mean–std normalization, suggesting that the super-resolution design of SRSNet, which relies heavily on spatial detail preservation, is more robust when the intensity distribution is constrained within a fixed and predictable range. This sensitivity underscores the importance of harmonizing preprocessing pipelines when deploying high-resolution segmentation networks in heterogeneous clinical environments.
A second aspect emerging from the results concerns the role of contextual information. While the introduction of additional slices markedly improves tumour segmentation performance in Attention-Res-U-Net, producing a clear monotonic gain as the spatial context increases, the same trend does not manifest for SRSNet. The Attention-Res-U-Net architecture benefits substantially from the richer 3D anatomical context, likely because attention mechanisms explicitly exploit long-range spatial relationships that allow the network to discriminate subtle tumour boundaries from surrounding parenchyma. SRSNet, by contrast, appears to saturate more quickly, indicating that its architecture already embeds strong local detail extraction capabilities and may not profit from wider anatomical fields, particularly when the target structure is extremely small relative to the input volume.
Data augmentation also plays a role in shaping model performance. For SRSNet, moderate

geometric transformations help improve tumour segmentation by introducing controlled variability, while more aggressive intensity augmentations amplify performance for the tumour but do not translate into improved pancreas delineation. This divergence suggests that large organs with consistent shape priors benefit from stable intensity profiles, whereas very small lesions profit from stronger perturbations that mimic the variability encountered in clinical practice. In contrast, the Attention-Res-U-Net configurations employ a consistent augmentation scheme across all tested variants, and the performance trends indicate that augmentation acts as a secondary factor; the quantity of contextual slices remains the dominant determinant of accuracy.

These observations collectively indicate that no single architecture is uniformly optimal across all anatomical targets. Super-resolution networks such as SRSNet demonstrate strong capability for large-structure segmentation, where spatial continuity and detailed boundary preservation are essential, whereas attention-based architectures excel in the highly challenging task of identifying small, low-contrast lesions. The interplay between normalization, data augmentation, and contextual depth further emphasises that segmentation performance cannot be attributed solely to architectural complexity but instead arises from a delicate balance between model design and data preparation.

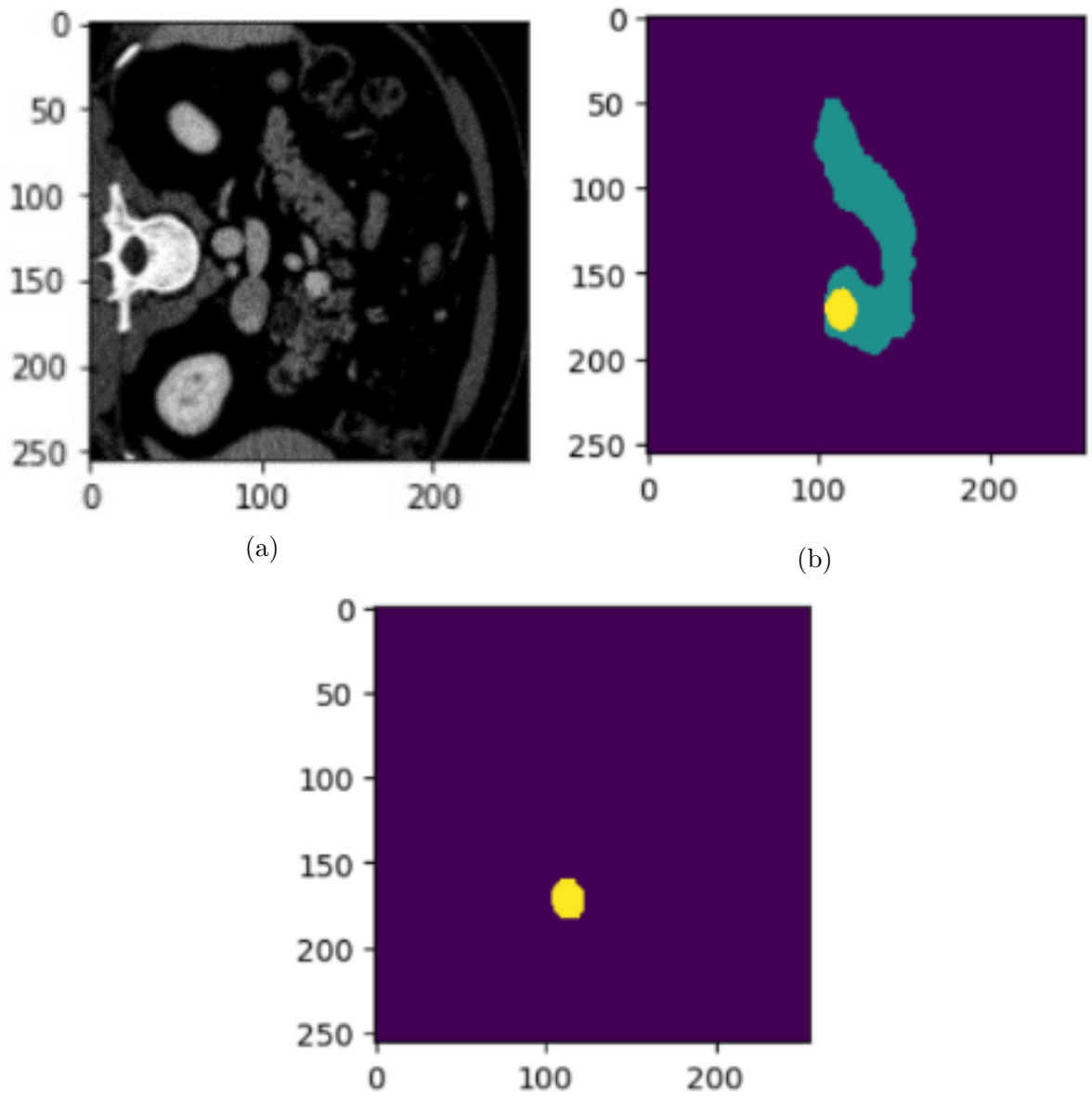
Overall, the results reinforce the value of adopting task-specific segmentation strategies within a broader clinical workflow. When integrated thoughtfully, these models can complement each other, leveraging their respective strengths to provide robust delineations for both organs at risk and tumour regions. Such insights contribute meaningfully to the translation of DL methods into routine medical imaging practice, where consistency, reliability, and adaptability are critical for supporting radiologists and improving downstream therapeutic planning.

In Fig. 3.6 is shown an example of tumour prediction on MSD image. Subsequently, SRSNet and Attention-Res-U-Net are tested on the CPTAC-PDA CT images to assess theirs ability to generalize and work with images and patients completely independent from those used for training. Since annotations are only provided for the tumor class, it is possible to compute the mean Dice score for this class only, which results in 42.1% for SRSNet and 43.2% for Attention-Res-U-Net. This result demonstrates the very good generalization properties of the DCNN model. One example of segmentation obtained on the CPTAC-PD dataset is illustrated in Fig. 3.7, depicting the tumor segmentation contours outlined by the expert radiologist in green, and the segmentation performed by model SRSNet represented in red for tumors and blue for the pancreas. The models

Table 3.1: Summary of segmentation performance across different neural network architectures (MSCA, MDAG, SRSNet, Attention-Res-U-Net). The table details the preprocessing configurations, including windowing, intensity normalization methods, data augmentation strategies (e.g., shift), and the resulting Dice Similarity Coefficient (DSC) for both single-class and multi-class target formulations.

Group	Model Name	Classes	Window	Normalization	Escl. bg	Epochs	Resampling			Slice in	Loss	Cl. Weig.	Aug.	Dice score			
							x	y	z					Per Slice		Per Pat.	
														T	P	T	P
MSCA	MSCA_v00	Tumour	[0, 255]	In 0 and 1 per CT image	NO	150	1 mm	1 mm	1 mm	1	C.E.	-	NO	93.20%	NA	¯4%	NA
	MSCA_v01	Tumour	NO	In 0 and 1 per CT image	NO	150	1 mm	1 mm	1 mm	1	C.E.	-	NO	86.70%	NA	¯4%	NA
	MSCA_v02	Tumour, Pancreas	NO	mean / std per CT image	NO	200	1 mm	1 mm	NO	1	C.E.	-	NO	-	-	4.10%	38.20%
	MSCA_v03	Tumour, Pancreas	[-100, 240]	mean / std per CT image	SI	200	1 mm	1 mm	NO	5	C.E.	-	rot + shift	-	-	0.40%	46.80%
	MSCA_v03_1	Pancreas	[-100, 240]	mean / std per CT image	SI	200	1 mm	1 mm	NO	5	C.E.	-	rot + shift	-	-	NA	51.10%
	MSCA_v04	Tumour, Pancreas	NO	mean / std per CT image	SI	200	1 mm	1 mm	1 mm	5	C.E.	-	NO	-	-	6.10%	29.20%
	MSCA_v05	Tumour, Pancreas	NO	mean / std per CT image	SI	200	1 mm	1 mm	NO	1	C.E.	-	rot + shift	-	-	2e-05%	10.60%
	MSCA_v06	Tumour, Pancreas	NO	mean / std per CT image	SI	200	1 mm	1 mm	1 mm	5	C.E.	-	shift	-	-	3.80%	47.70%
	MSCA_v07	Tumour, Pancreas	NO	mean / std per CT image	SI	200	1 mm	1 mm	1 mm	5	C.E.	-	bright	-	-	3.30%	26.30%
MDAG	MDAG_v01	Tumour	[0, 255]	In 0 and 1 per CT image	NO	200	1 mm	1 mm	1 mm	1	C.E.	-	NO	-	-	8.50%	NA
	MDAG_v02	Tumour	NO	mean / std per CT image	NO	200	1 mm	1 mm	1 mm	1	C.E.	-	NO	-	-	4.00%	NA
	MDAG_v03	Tumour, Pancreas	NO	mean / std per CT image	NO	200	1 mm	1 mm	1 mm	1	C.E.	-	NO	-	-	15.40%	53.50%
SRSNet	SRSNet_v01	Tumour, Pancreas	[-100, 240]	In 0 and 1 per CT image	SI	200	1 mm	1 mm	NO	1	C.E.	-	NO	-	-	17.70%	64.00%
	SRSNet_v02	Tumour, Pancreas	[-100, 240]	mean / std per CT image	SI	50	1 mm	1 mm	NO	1	C.E.	[1,200,1200]	NO	-	-	12.20%	44.90%
	SRSNet_v03	Tumour, Pancreas	[-100, 240]	mean / std per CT image	SI	50	1 mm	1 mm	NO	1	C.E.	-	rot + shift	-	-	16.20%	66.80%
	SRSNet_v04	Tumour, Pancreas	[-100, 240]	mean / std per CT image	SI	50	1 mm	1 mm	1 mm	11	C.E.	-	NO	-	-	18.40%	65.80%
	SRSNet_v05	Tumour, Pancreas	[-100, 240]	mean / std per CT image	SI	50	1 mm	1 mm	1 mm	11	C.E.	-	rot + shift + bright	-	-	37.50%	65.10%
Attention-Res-U-Net	AUnet_v01	Tumour	[-100, 240]	In 0 and 1 per CT image	SI	200	1 mm	1 mm	1 mm	1	Jaccard.	-	rot + shift + noise	-	-	36.50%	NA
	AUnet_v02	Tumour	[-100, 240]	In 0 and 1 per CT image	SI	200	1 mm	1 mm	1 mm	5	Jaccard	-	rot + shift + noise	-	-	41.20%	NA
	AUnet_v03	Tumour	[-100, 240]	In 0 and 1 per CT image	SI	200	1 mm	1 mm	1 mm	11	Jaccard	-	rot + shift + noise	-	-	45.40%	NA

and the training and test script are written using the Tensorflow framework. The model Attention-Res-U-Net is developed, trained and tested in PyTorch.



(a) (b)

(c)

Figure 3.6: The result of prediction with SRSNet model on MSD image: (a) CT image of pancreas and tumour; (b) Label image of pancreas and tumour; (c) Predicted tumour.

.

3.2 Two step segmentation of pancreas

The previously reported results were obtained using a one-step segmentation process. The segmentation framework presented here incorporates an initial step designed to localize a smaller region of interest within the full CT scan that contains the tumor, followed by segmentation performed exclusively on this volume. Two methods are developed for

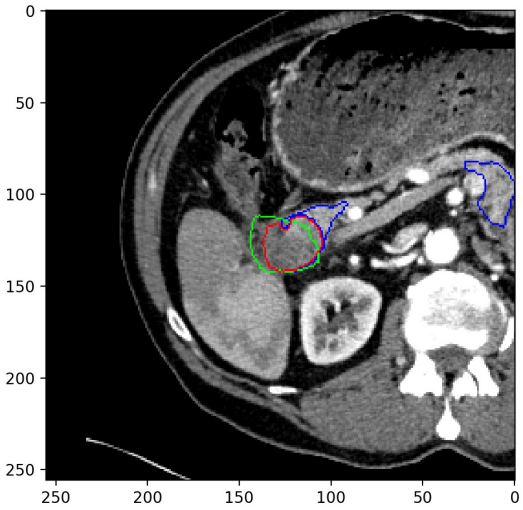


Figure 3.7: Contour on label and prediction for tumour class.

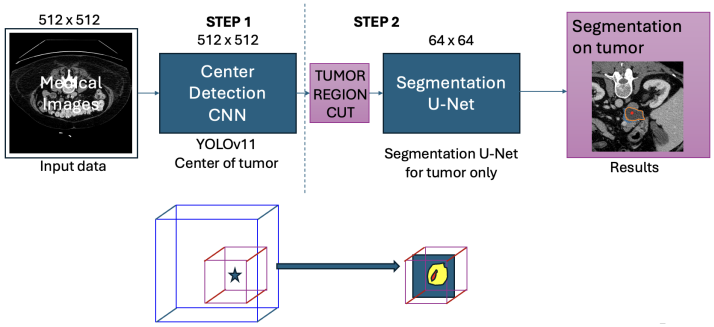


Figure 3.8: The pipe-line of the study, from the localization of the model, to the cropped image and finally to segmentated tumor.

tumor localization and volume-of-interest cropping:

- Divide the full CT image into small size cubes and train a classification model to discriminate if there is the tumor of pancreas. In the second step take the predicted cubes and perform segmentation over all the slides.
- Train an object detection model to find the pancreas or the tumor of pancreas into the 3D image. In the second step crop the size of a cube centered in the geometrical center of the detected pancreas and perform segmentation over all the slides.

The two methods are developed in the Tensorflow framework. The two pipeline are developed independently.

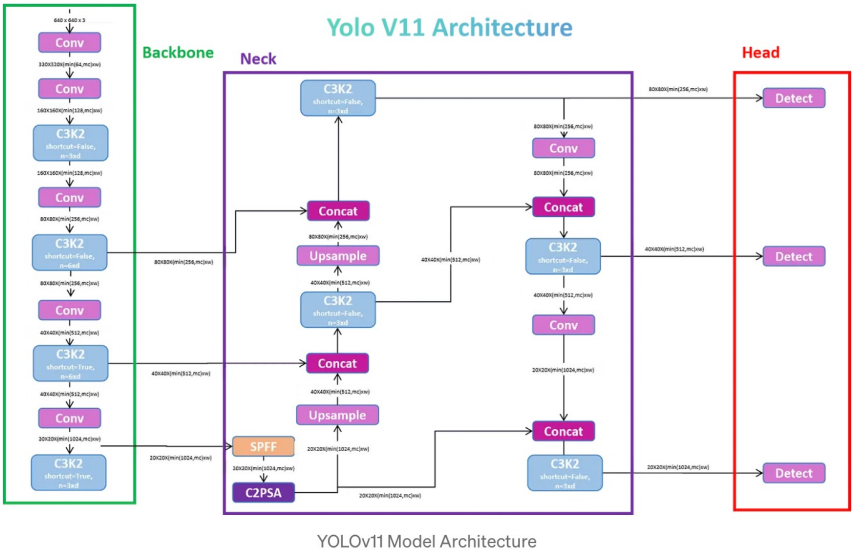


Figure 3.9: Model representation of the YOLO v11 architecture used for the initial coarse localization of the pancreas and tumor bounding boxes in Step 1B.

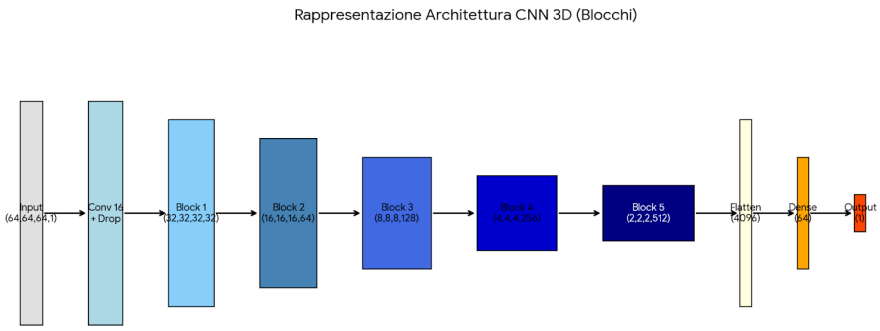


Figure 3.10: Architecture of the CNN 3D classificator model utilized in Step 1A to identify patched volumes containing the pancreas or pancreatic tumors.

A two-step pipeline has proven effective:

1. **Step 1A**: Coarse localization of the pancreas through a 3D-CNN classificator model that find the patched volumes of CT images with pancreas or tumour of pancreas. Fig. 3.10 shows the model used for classify the patched volume.

2. **Step 1B**: Coarse localization of the pancreas through a YOLO (v11) model that find the bounding boxes of pancreas and tumour. Fig. 3.9 shows the logical architecture of the YOLO v11 model for object detection.

3. **Step 2**: Fine segmentation using a high-resolution network such as 2D U-Net, that takes the patched volume and segment the image slice per slice.

The following sections describe each step. The Fig. 3.8 could summarize the pipe-line of the study.

3.2.1 Step 1: Pancreas Localization

The step 1 of the localization task is performed in two ways in this study and the two performance are evaluated separately.

Step 1A: Classification of patches Because the main result is to roughly localize the tumour of pancreas, the idea is to use the same label for pancreas and tumour, as a logic consequence of the fact that tumour affects the pancreas tissue. In order to overcome the inbalance issue, the dataset is equally made randomly taking the 50% of the patches on the background class, and the other 50% of the patches randomly picked-up on the class pancreas-tumour. The dataset was divided by 75% for the training and 25% for the validation. The loss used for the training and validation is the Cross Entropy, the number of epochs is 50. Classification metric used is the F1-Score:

$$\mathrm{F1} = \frac{2TP}{2TP + FP + FN} \tag{3.2}$$

that is the semplification of the combined formula:

$$\mathrm{F1} = \frac{Precision \cdot Recall}{Precision + Recall} \tag{3.3}$$

where

$$\mathrm{Precision} = \frac{TP}{TP + FP} \tag{3.4}$$

and

$$\mathrm{Recall} = \frac{TP}{TP + FN} \tag{3.5}$$

where the used symbols are:

1. TP stands for True Positive outcomes,
2. FP stands for False Positive outcomes,
3. FN stands for False Negative outcomes

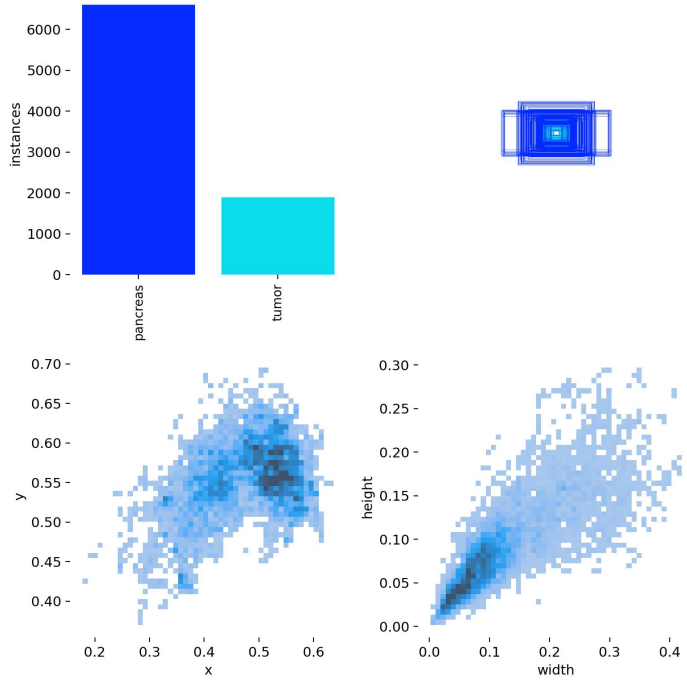


Figure 3.11: Statistics of the bounding boxes, on the top left the number of bounding boxes per class type. On the top right the different sizes of the bounding boxes. On the bottom left a correlation plot for x and y, to see the distribution of the bounding boxes in the image. On the bottom right a correlation plot for *width* and *height* values, as is possible to see the distribution is concentrated for small values.

Model	epochs	train loss	val loss	train F1	val F1
3D Unet	50	0.0016	0.015	0.44	0.34

Table 3.2: Summary of the classification task results in Step 1A, detailing epochs, training/validation loss, and F1 scores.

The F1 is the score that comes from the result for the training and validation of the model are summarized in Tab. 3.2.

Step 1B: Bounding Box Regression YOLO v11 The used architecture for localization is a 3D CNN, the architecture YOLO, the 11 version of the structure, that outputs a bounding box:

$$(x, y, z, w, h, d) \tag{3.6}$$

representing the center and dimensions of the predicted pancreas region. The goal is not pixel precision, but to tightly enclose the organ for the segmentation network. The dataset is pre-processed entirely to extract the bounding boxes from the labels, where

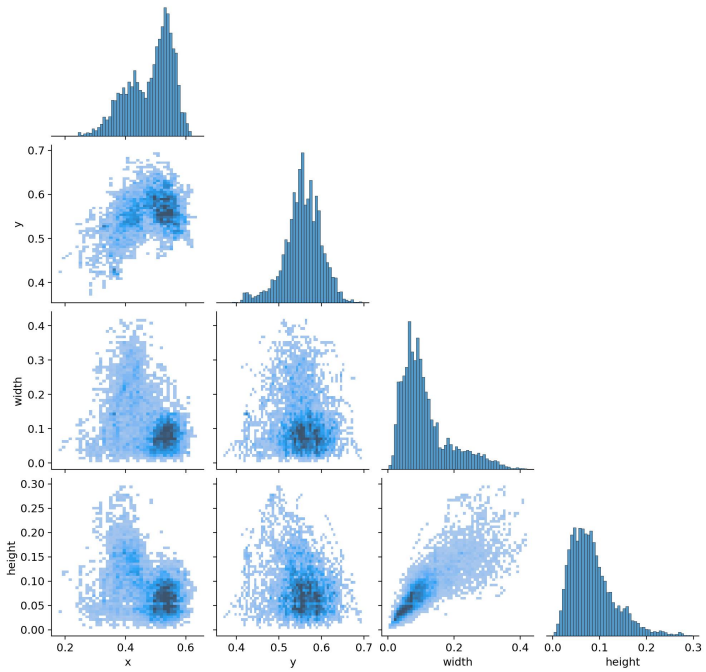


Figure 3.12: Correlogram plot is the correlation plot between the distribution of x, y, *width* and *height* values. In the diagonal of the correlation matrix there is distribution of all the 4 variables.

the pancreas and the tumour are fully inside the edges. The new dataset consists of images and bounding boxes for every image. The multi-channel option is used: the dataset contains the bounding boxes of the pancreas and the bounding boxes of the tumour. Fig. 3.11 summarized the dataset values of x, y, the normalized coordinates of the center of the boxes, and *width* and *height*, the normalized sizes of the boxes. While Fig. 3.12 put in correlation the 4 values.

Training and Metrics results The localization model is trained using a regression loss such as:

$$\mathcal{L} = \left\| \hat{b} - b \right\|_1 \tag{3.7}$$

where $\hat{b}$ is the predicted bounding box and b is the ground truth. The dataset is splitted in 75% of training and 25% of validation, as the Step 1A. The number of training patients is 210, while the number of validation patients is 71. The training procedure last 100 epochs, and the best results are obtained at 50 epochs. In order to see that, Fig. 3.13 shows the losses values and the metrics values at every epochs. The results of the confusion matrix

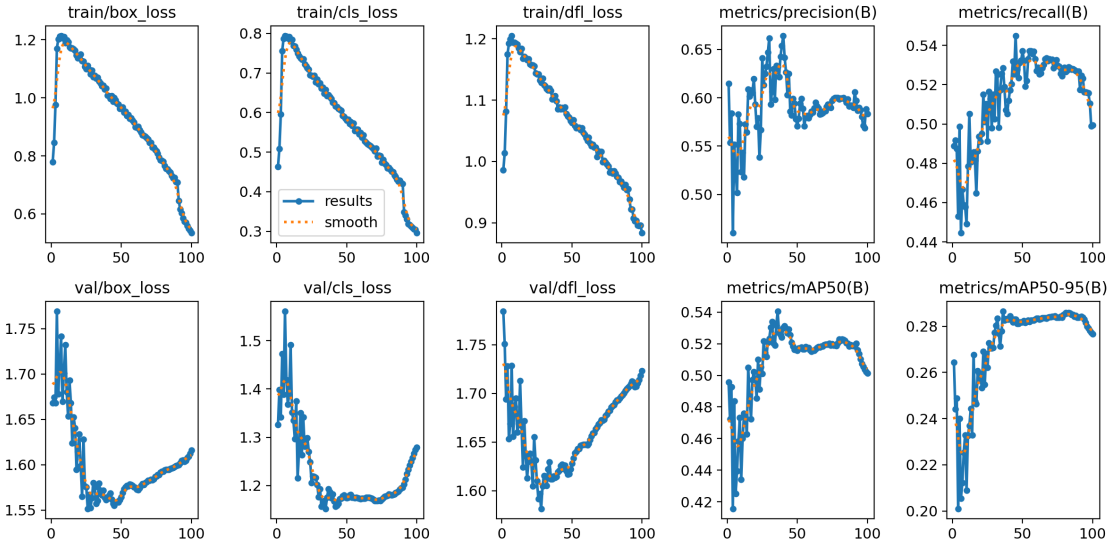


Figure 3.13: Losses and metrics: the "$train/box_{loss}$" is the training loss on the bounding box object, a loss that measures how "tight" the predicted bounding boxes are to the ground truth object. The "$train/cls_{loss}$" is the training loss on the cluster box, a loss that measures the correctness of the classification of each predicted bounding box: each box may contain an object class, or a "background". This loss is usually called cross entropy loss. The "$train/dfl_{loss}$" is the distribution focal loss during the training, a loss that optimizes the coordinates regression as probability distributions. The same values are calculated for the validation set, at the bottom of the image. The metrics Precision and Recall are shown at the top, while at the bottom are shown the mean Average Precision at 50 level of IoU and between 50 and 95 of IoU.

are shown in Fig. 3.14. The results are fully summarized by the correlated metrics plots, in Fig. 3.15.

Some examples illustrating the real images with training and predicted bounding boxes are summarized in Fig. 3.16 and in Fig. 3.17.

ROI and center extraction Once the predicted bounding box is available, the CT volume is cropped around the pancreas, significantly reducing the input domain of the segmentation model. The crop is made around the center. The crop is 64 × 64 large in pixels. The center is calculated once the predicted tumour is reconstructed in all the ROI of the CT image. The centroid of the labels predicted represent the center of tumour and pancreas. In that way, if the tumour is not predicted at all, the center of the pancreas is used as center for the cropping operation. In Fig. 3.18 two examples of center correctly found. The distance between the centers of labelled tumour and the predicted one is

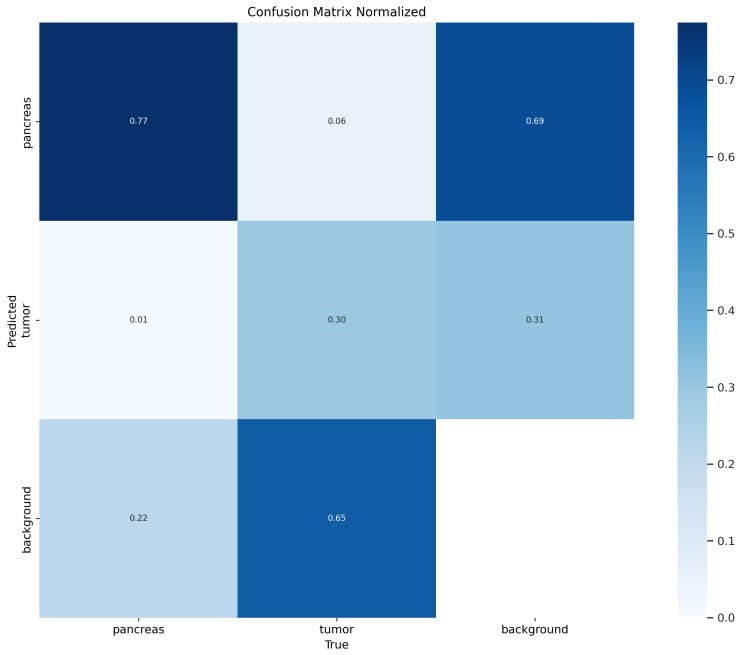


Figure 3.14: The confusion matrix, the results of the object detection summarized.

used at metrics to evaluate the process. In Fig. 3.19 are summurized the plots of the distance in 3 dimensions and the distance on z-axis.

3.2.2 Step 2: Segmentation of Tumour

After isolating the pancreas region, the second stage performs fine segmentation of the PDA tumour. The segmentation now is performed over the tumour alone, on a smaller region of fixed size 64 $\times$ 64. It is performed slice per slice, but taking into account a number of channel per image 1 and 11, for the 2.5D model test. The training and test split is exactly the same used for the training and test of the object detection model, in order to evaluate the entire process. But in order to augment the number of images seen by the Unet, and balance the dataset, an equal number of image of the background, selected where there is no labels coming from pancreas also, and the same number of images of tumour, are selected randomly. The augmentation than follows with shift of the images on the center.

Model U-Net The used architecture is a Unet with 2D convolution, which was previously used for segmentation of pancreas in early stage study. This simple model performs the single slice segmentation in one-channel and multi-channel mode.

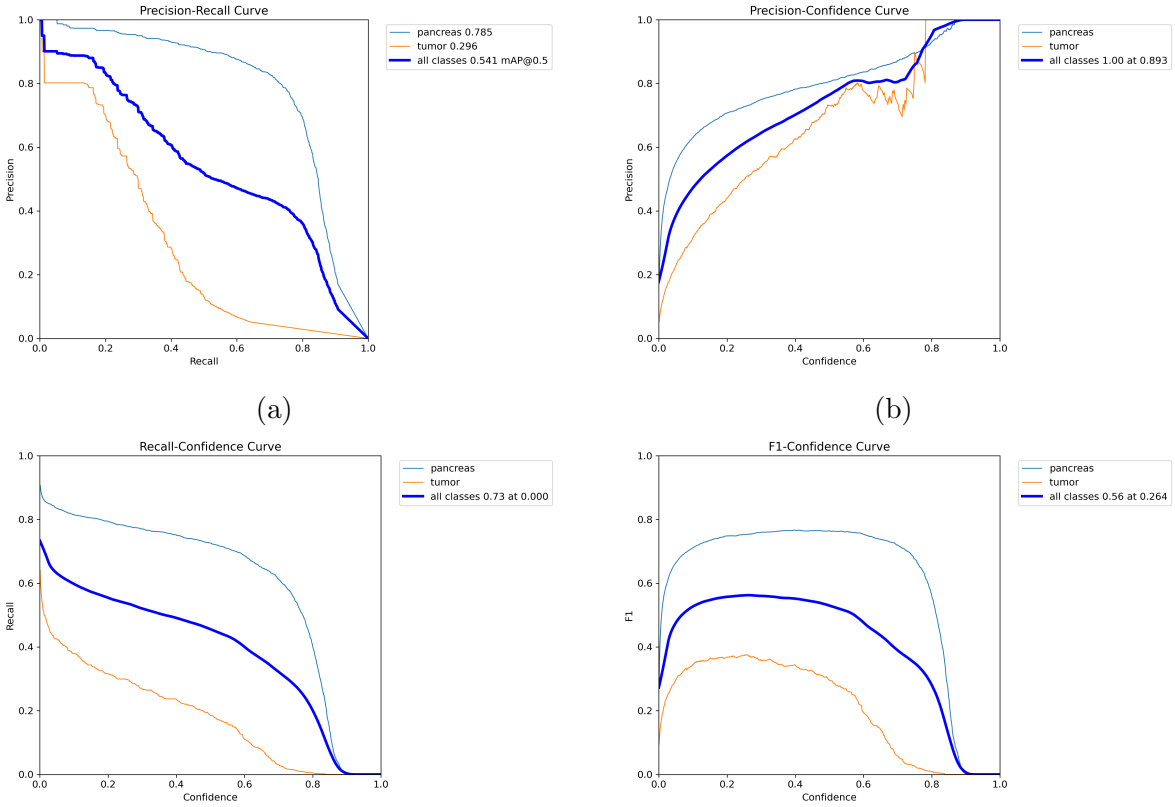


Figure 3.15: (a)PR plot is the correlation plot between Precision and Recall, it is calculated for the mAP@50 value 0.541, the mean of the area under the curve for all the classes. The prediction of the pancreas is good, because the curve is high and squared, while the tumor curve goes down early, so there are a lot of FP.(b)P plot shows the value of the Precision changing in function of the confidence level to accept the prediction. The curve is very low for low value of the coincidence, because the false positive prediction number is high. In middle values the curve rising up. Near confidence 0.7-0.8 the value of precision remains stable, because of rising False Negative. The curve remains high until 1.(c)R plot shows the value of the Recall changing in function of the confidence level. It is the opposite plot of Precision curve. The curve is high with low confidence, because of high number of false positives. In the middle the recall goes down due to lose of false positive togheter with some true positive. The curve goes to zero, because at high level of confidence, many objects are not detected.(d)F1 plot shows the value of the F1 score changing in function of the confidence level. It is the value of intersection between the R and P curve. It is low for lo confidence, because of high number of FP. It is maximum in the middle for a balance between R and P values. It goes down because high number of FN. For pancreas class confidence is good at 0.43, but for tumour is lower, near 0.26.

Loss Functions Because the PDA volume is often extremely small relative to the pancreas and background, a specialized loss functions help mitigate class imbalance: the

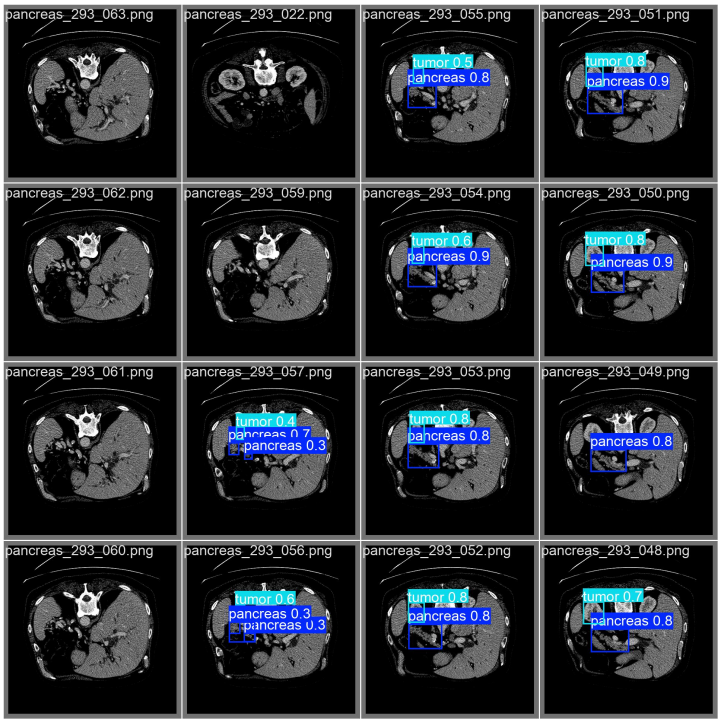


Figure 3.16: Image of validation set as example, predicted classes for pancreas and tumour.

Focal Cross Entropy loss for tumor, with an $\alpha = 0.72$ and a $\gamma = 1.6$, two values reported in state-of-art.

Results of Step 1A and 1B The Step 1A metric used is the validation F1 score (Tab. 3.2). The score is low in classification of pancreas tumor state-of-art [98, 99]. The Step 1B metric used are the center's distance of tumour (Fig. 3.19) and the F1 score in Fig. 3.15. As shown the low value of F1 score for tumour (Fig. 3.15d) is not a issue, because is balanced by the high value of F1 score on pancreas. As said the crop is made on tumour and pancreas center. The conclusion is to follow the Step 1B, but the results will be shown for all the two pipe-line.

Results of Step 2 The results of the metric for segmentation DSC is:

- DSC for 2 D model: 0.66
- DSC for 2.5 D model: 0.62

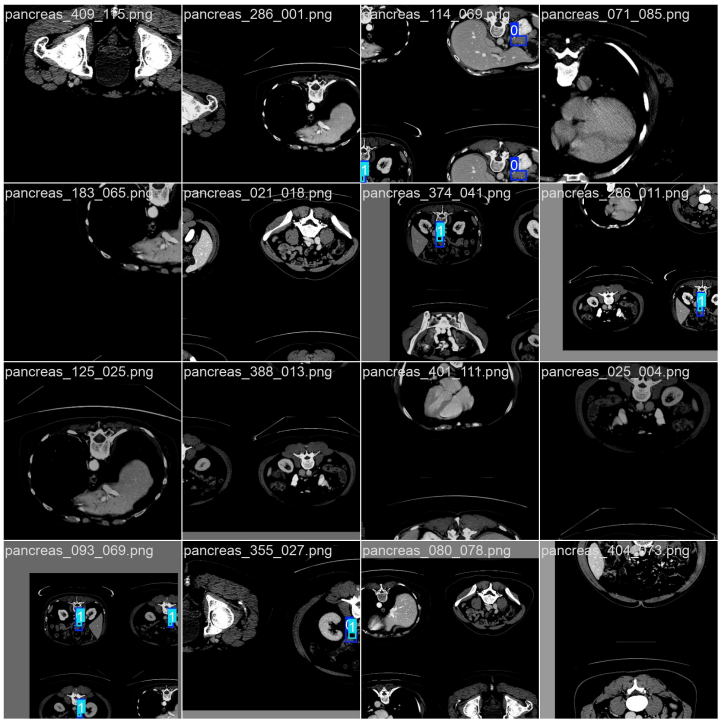
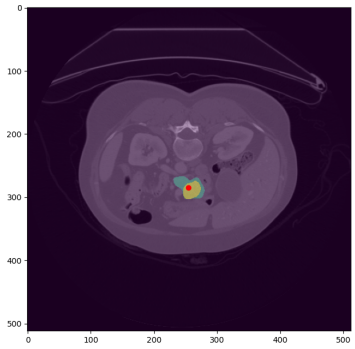


Figure 3.17: Image of training set example, the augmentation technique of YOLO with cropping, patching sequence, rotation and traslation.

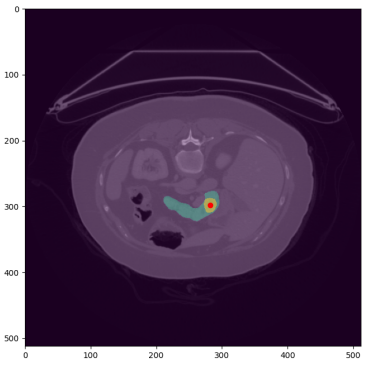
that shows a better result for a simple 2D (Fig. 3.20), opposite to clear conclusion on SRSNet model performance in One Step analysis, due to the difference on image dimesions. The cropping operation increases the number of labelled tumour pixels and decreases the background.

3.2.3 Results and Discussion

The full pipe-line is now possible to be addressed. Taking the same dataset division in training and validation used for the two step, the model is feeded with validation set, the centers obtained are used for the crop process, the patched volume is given to the segmentation model, the DCS is then calculated over the full validation set. The result is showed in Tab. 3.3 where are also reported the DCS from a state-of-art paper, the two DCS of the One Step study, proving clearly that the Two Step pipeline offers clear advantages: dramatically reduces the search space for segmentation, improves segmentation accuracy by focusing the network on the correct anatomical region, allows separate optimization of localization and segmentation tasks, is robust to anatomical variability and differences in CT coverage.
The Two Step pipe-line with Step 1A shows a dramatically reduction of the full DSC due



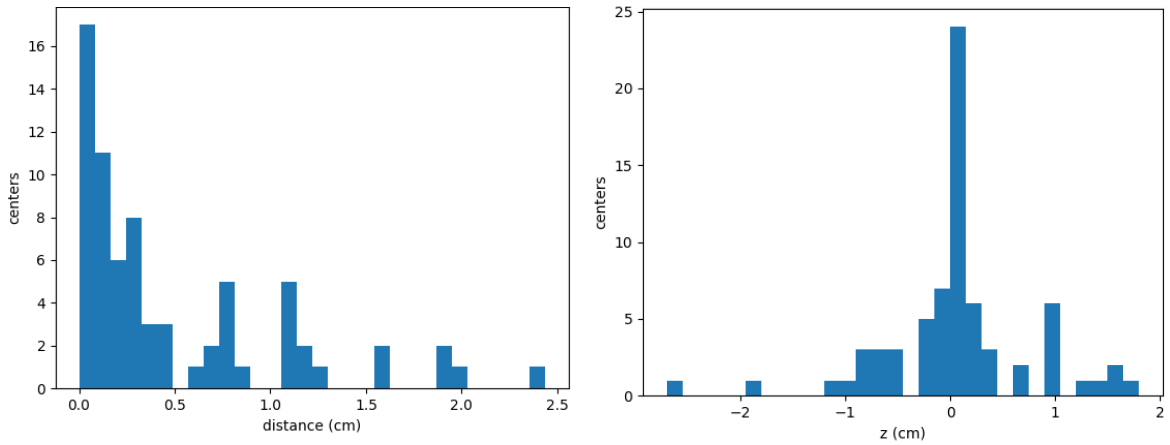
(a)



(b)

Figure 3.18: Two example of tumour center that is correctly found (a-b). The pancreas is blue zone, while the tumour is yellow zone. A crop on the center of pancreas in image (b) might include the tumour anyway.

to two main problems in the process: one that, although the F1 score for the classification in Step 1A and the F1 score in the object detection of Step 1B are quite the same, the localization of Step 1B happens in the bounding boxes, that entirely localize the tumour. And second, the Step 1B center extraction includes the pancreas volume, where the tumour is generally localized.



(a) (b)

Figure 3.19: (a) Histograms of the distance in cm between the true center and the predicted center. (b) Histogram of the z-distance in cm between true and predicted center.

Pipe-line	DSC tumour (%)
One Step SRSNet [100]	37.50%
One Step Attention-Res-U-Net	45.40%
Two Step - 1A+2	22.30%
Two Step - 1B+2	**58.00**%
Two Step State-of-art [95, 101]	54.00%

Table 3.3: Comparative analysis of tumor segmentation accuracy (DSC) between One-Step and Two-Step pipelines, demonstrating the significant improvement achieved by prior localization.

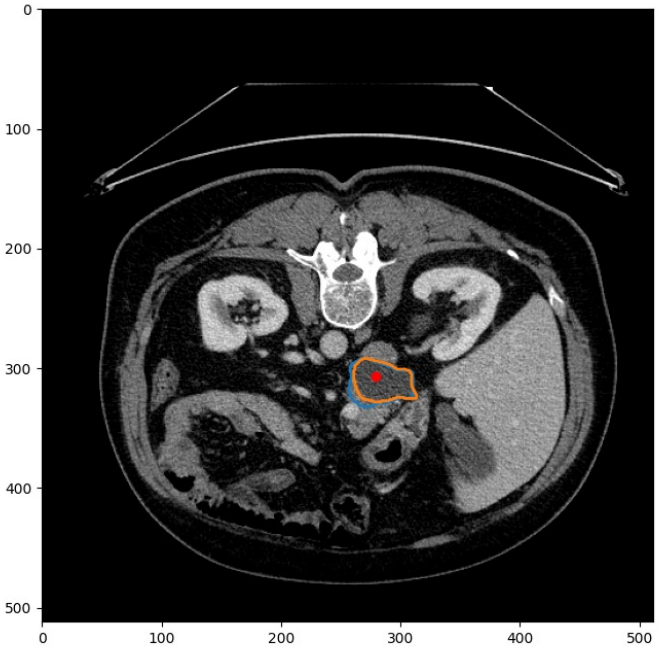


Figure 3.20: Image of a validation slice correctly centered and with a DSC=87%. The orange line contours the true tumour, while blue line contours the predicted one.

Chapter 4

SPECT reconstruction algorithm in innovative neutron therapy

Boron Neutron Capture Therapy (BNCT) is an innovative and highly targeted form of radiotherapy that holds significant promise in the treatment of various cancers [102], particularly those that are difficult to treat with conventional methods. Boron Neutron Capture Therapy (BNCT) is a biologically targeted form of radiotherapy based on the nuclear reaction between boron-10 and low-energy neutrons, producing high-linear-energy-transfer particles that selectively destroy tumor cells while largely sparing surrounding healthy tissue. Clinical experience to date indicates that BNCT has shown the most consistent therapeutic success in recurrent or locally advanced head and neck cancers, particularly squamous cell carcinomas that have failed prior surgery, external-beam radiotherapy, and chemotherapy. In these patients, BNCT has achieved meaningful local control and survival outcomes where conventional salvage options are limited by toxicity.

BNCT has also been extensively studied in malignant gliomas, including glioblastoma multiforme. In this setting, BNCT has been used as an adjunct or salvage treatment following maximal safe surgical resection and standard chemoradiotherapy with temozolomide. Although overall prognosis remains poor, BNCT has demonstrated improved local tumor control and modest survival benefits compared with historical outcomes using conventional therapy alone. Additional clinical success has been reported in selected cases of cutaneous melanoma and extramammary Paget's disease, where the superficial nature of the tumors and favorable boron uptake have led to high response rates with acceptable toxicity and good cosmetic outcomes.

Early-phase clinical studies have explored BNCT in hepatocellular carcinoma, recurrent lung tumors, sarcomas, and other rare malignancies, particularly in patients who are not candidates for further surgery or radiation. While these results are promising, evidence remains limited and further trials are required to establish efficacy relative to standard treatments.

Conventional management of the cancers treated with BNCT typically relies on established multimodal approaches. Head and neck cancers are primarily managed with surgery, intensity-modulated radiotherapy, chemotherapy, targeted agents, and immunotherapy. Malignant gliomas are treated with surgical resection followed by fractionated radiotherapy and alkylating chemotherapy. Melanoma is usually treated with wide surgical excision and systemic immunotherapy or targeted therapy, with radiotherapy reserved for selected cases. Hepatocellular carcinoma is managed using liver resection, ablation, transarterial therapies, systemic targeted or immune-based drugs, and liver transplantation in eligible patients. In most clinical settings, BNCT remains an investigational or complementary modality, typically reserved for recurrent, radioresistant, or otherwise treatment-refractory disease. In the context of BNCT, estimating the dose delivered to the patient using a tomography set-up with gamma cameras is a major long-term clinical goal. The work presented in this framework represents a fundamental step toward that goal by investigating image reconstruction techniques on controlled experimental setups.

4.1 Neutron therapy highlights

The fundamental principle of BNCT lies in its two-step approach: first, the boron-10 (^{10}B) containing compound is selectively delivered to tumor cells. This compound is designed to accumulate preferentially in cancer cells due to their altered metabolism and increased uptake rates. [103]. The second step consist of the neutron irradiation: once the boron compound has concentrated in the tumor, the patient is exposed to a beam of low-energy (thermal) neutrons. These neutrons interact with the boron-10 atoms, initiating a nuclear reaction [104]. The nuclear reaction that occurs is the core of BNCT's effectiveness:

$$^{10}\mathrm{B} + {}^{1}\mathrm{n} \longrightarrow {}^{7}\mathrm{Li} + {}^{4}\mathrm{He} + 2.79\,\mathrm{MeV} \tag{4.1}$$

This reaction produces high-energy alpha particles (^{4}He) and lithium nuclei (^{7}Li). These particles have a very short range in tissue (5-9 μm), which is approximately the

diameter of a single cell. This limited range ensures that the destructive energy is confined primarily to cells that contains the boron compound, providing a highly localized and targeted treatment.

The therapeutic efficacy of BNCT is governed by a combination of tightly coupled physical, chemical, and biological factors, among which the most critical is the boron delivery agent and its ability to achieve highly selective accumulation of boron-10 within malignant cells while minimizing uptake in normal tissues. Effective BNCT requires intracellular boron concentrations on the order of 20–40 μg of ^{10}B per gram of tumor tissue, together with tumor-to-normal tissue and tumor-to-blood concentration ratios typically exceeding 3:1. To date, clinical applications of BNCT have relied primarily on two boron-containing compounds: boronophenylalanine (BPA) and sodium borocaptate (BSH). BPA is an amino-acid analogue that exploits the upregulation of L-type amino acid transporters (LAT1) commonly observed in rapidly proliferating cancer cells, enabling preferential uptake via active transport mechanisms and partial intracellular localization. In contrast, BSH is a boron-rich dodecaborate compound that does not readily cross intact cell membranes and therefore accumulates predominantly in the extracellular and vascular compartments, making its effectiveness more dependent on tumor vascular permeability and blood–tumor barrier disruption. Ongoing research is directed toward next-generation boron carriers, including boronated peptides, antibodies, liposomes, nanoparticles, and porphyrin-based compounds, with the aim of improving tumor selectivity, intracellular targeting, and retention time.

A second essential determinant of BNCT effectiveness is the neutron source, whose performance is defined primarily by neutron energy spectrum, flux intensity, beam collimation, and contamination from fast neutrons and gamma radiation [105, 106]. An optimal neutron beam must deliver epithermal neutrons with sufficient intensity to penetrate tissue and thermalize at tumor depth, while minimizing dose to surrounding healthy structures. Historically, BNCT was performed using research nuclear reactors, which provided high neutron fluxes but posed significant logistical, regulatory, and accessibility challenges for routine clinical use. More recently, accelerator-based neutron sources have emerged as a practical alternative, offering hospital-based installation, improved safety profiles, and greater flexibility in beam shaping and treatment planning [107, 108]. The convergence of advances in boron pharmacology and accelerator-driven neutron technology is central to the continued clinical translation and optimization of

BNCT. [107, 108]. The third is the dosimetry: accurate measurement and calculation of the radiation dose delivered to tumor and healthy tissues are essential for treatment planning and evaluation. This is the focus of this thesis. The fourth is the tumor biology, because understanding the biological characteristics of different tumor types is crucial for optimizing boron delivery and treatment efficacy[109]. Boron Neutron Capture Therapy (BNCT) offers several significant advantages that distinguish it from conventional radiotherapy techniques, while also presenting notable technical and clinical challenges.
One of the primary strengths of BNCT lies in its high selectivity. The therapy relies on the preferential accumulation of boron-containing compounds within tumor cells, followed by a localized nuclear reaction triggered by neutron irradiation. This reaction occurs at the cellular level, allowing for highly precise targeting of malignant tissue while sparing nearby healthy cells.

Closely related to this selectivity is the reduction in side effects. The high linear energy transfer (LET) particles produced during the boron–neutron capture reaction have a very short path length, typically limited to the dimensions of a single cell. As a result, radiation damage is largely confined to boron-loaded tumor cells, minimizing collateral damage to surrounding normal tissues and potentially improving patient quality of life during and after treatment.

Another important advantage of BNCT is its potential effectiveness in treating deep-seated tumors. Unlike charged particles or photons used in conventional radiotherapy, neutrons possess a high penetration capability, enabling them to reach tumors located deep within the body. This characteristic makes BNCT a promising option for malignancies that are difficult to access or inadequately treated using standard radiation techniques.

In addition, BNCT may allow for treatment to be delivered in a single session in certain clinical scenarios. Because a high radiation dose can be deposited selectively within tumor cells during one irradiation, the need for multiple treatment sessions may be reduced. This can lower the overall treatment burden for patients and decrease the number of required hospital visits.

Despite these advantages, several challenges limit the widespread clinical adoption of BNCT. One major obstacle is the requirement for complex infrastructure. The therapy depends on specialized neutron sources, such as nuclear reactors or accelerator-based systems, along with heavily shielded treatment facilities. These requirements significantly

increase costs and restrict availability to a small number of centers worldwide.

Another critical challenge is the optimization of boron delivery. Achieving a sufficiently high and selective concentration of boron within tumor cells, while maintaining low uptake in healthy tissues, remains an active area of research. The effectiveness and safety of BNCT are strongly dependent on the development of improved boron carriers with favorable biological and pharmacokinetic properties.

Finally, treatment planning for BNCT is particularly complex due to the interplay of neutron physics, radiation transport, and radiobiological effects. Accurate dose calculation requires sophisticated treatment planning systems capable of modeling multiple radiation components and their biological impact. Continued advancements in computational methods and dosimetry are therefore essential for the safe and effective clinical implementation of BNCT.

BNCT has shown promising results in treating various cancers, including glioblastoma multiforme, recurrent head and neck cancers, and malignant melanomas. As research progresses and clinical trials expand, BNCT has the potential to become a valuable addition to the arsenal of cancer treatments, offering hope for patients with otherwise difficult-to-treat malignancies.

4.1.1 Accelerator based BNCT facilities

The way to produce the neutron beam with the best energy to produce the reaction with Boron compound is essentially one: to accelerate charged particle on a target, then modify the kinetic energy of the produced neutrons to match the low values that are suitable for the therapy. Accelerator based BNCT facilities (AB-BNCT) are compact facilities that generate neutrons through nuclear reactions induced by charged particles, typically protons or deuterons, accelerated to a few MeV of kinetic energy. The principle is to direct a high-current beam of charged particles onto a light-element target, where nuclear reactions produce neutrons with suitable energies and yields for BNCT [110, 111].

Several reactions have been explored for this purpose:

$$^{7}\mathrm{Li}(p,n)^{7}\mathrm{Be}, \quad E_{\mathrm{threshold}} = 1.88\,MeV, \tag{4.2}$$

$$^{9}\mathrm{Be}(p,n)^{9}\mathrm{B}, \quad E_{\mathrm{threshold}} = 2.06\,MeV, \tag{4.3}$$

$$^{3}\mathrm{H}(d,n)^{4}\mathrm{He}, \quad E_{\mathrm{threshold}} = 0.1\,MeV. \tag{4.4}$$

Among these, the $^{7}\mathrm{Li}(p,n)^{7}\mathrm{Be}$ reaction is the most widely adopted for accelerator-based BNCT because it can be driven by low-energy proton beams (around $2.5\,MeV$), resulting in relatively monoenergetic neutrons with energies peaked around $30\,keV$. The $^{9}\mathrm{Be}(p,n)^{9}\mathrm{B}$ reaction produces a broader neutron energy spectrum but requires slightly higher proton energies. The $^{3}\mathrm{H}(d,n)$ reaction, while efficient, involves tritium handling and associated radiological risks, making it less practical for medical environments.
The target must support high power densities due to the intense beam current (typically in the range of $5\,mA$ to $30\,mA$). Targets are often made of metallic lithium or lithium compounds deposited on high-thermal-conductivity substrates such as copper or molybdenum.
Active cooling systems, using either water or liquid metal heat exchangers, are essential to dissipate several kilowatts of beam power while maintaining target integrity and neutron yield stability.
The neutron yield and energy distribution strongly depend on beam energy. Operating just above the threshold energy ensures that most neutrons are emitted in the forward direction with lower average energies, facilitating subsequent moderation. Typical operating parameters for hospital-based BNCT accelerators are:

- Proton energy: $2.3 - 2.6\,MeV$
- Beam current: $5 - 10\,mA$
- Neutron yield: $\sim 10^{13} - 10^{14}$ n/s

The accelerator technology employed is based on large accelerators used for particle physics. Modern BNCT facilities employ compact proton accelerators such as radio-frequency quadrupoles (RFQs), drift tube linacs (DTLs), or cyclotrons. Linac-based systems are advantageous for their high beam quality and tunable energy, while cyclotrons offer simplicity and compactness. Shielding design is a critical aspect due to the coproduction of secondary gamma radiation and activation of structural materials.
Accelerator-based sources enable the installation of BNCT facilities in hospitals, eliminating the need for research reactors. They offer operational flexibility, on-demand beam delivery, and simpler regulatory management. Research efforts continue to optimize target materials, beam shaping assemblies, and accelerator performance to increase neutron flux while minimizing gamma contamination and maintenance requirements.

4.1.2 Neutron Moderation and Beam Shaping

The primary neutrons emitted are predominantly fast, with energies ranging from tens of keV up to several MeV, depending on the reaction and beam energy. However, effective BNCT treatment requires an epithermal neutron beam (neutron energies between $1\,eV$ and $10\,keV$). Neutrons in this energy range can penetrate several centimeters into tissue before being thermalized, allowing for selective boron capture in tumors located below the skin surface.

To transform the primary fast neutron spectrum into a clinically suitable one, a Beam Shaping Assembly (BSA) is used. The BSA is composed of several functional layers designed to moderate, reflect, filter, and collimate the neutron flux.
The moderator slows down fast neutrons through elastic scattering while minimizing absorption. Its purpose is to produce a smooth epithermal spectrum with minimal high-energy and thermal neutron components. Common moderator materials include:

- Aluminum fluoride (AlF_3)
- Magnesium fluoride (MgF_2)
- Teflon (PTFE)
- Light water or heavy water (in specific configurations)

The moderator's geometry and thickness are carefully optimized using Monte Carlo simulations (e.g., MCNP, PHITS, GEANT4) to achieve the desired neutron energy distribution and spatial uniformity.

Reflectors surround the moderator to reduce neutron leakage by scattering escaping neutrons back toward the beam axis. Materials such as graphite, beryllium, or lead are commonly used due to their high scattering cross-sections and relatively low absorption probabilities.

Filters are inserted to remove unwanted low-energy neutrons and gamma rays that contribute to patient dose without therapeutic benefit. Typical filter materials include sulfur, cadmium, or bismuth. Additional gamma shielding layers—often made of lead or tungsten—are implemented to attenuate prompt gamma rays generated in the target and moderator.

Finally, the beam is collimated to produce a well-defined, uniform neutron field at the treatment port. Collimators are made of high-hydrogen-content materials (such as polyethylene or borated polyethylene) to absorb scattered neutrons and reduce out-of-field dose. The exit aperture is shaped according to clinical requirements, typically circular or rectangular, depending on treatment field size.

The quality of the BNCT beam is characterized by several figures of merit:

- **Epithermal neutron flux:** typically $\sim 10^9$ n/cm^2/s at the beam port.
- **Contamination ratio:** the proportion of fast neutrons and gamma dose to epithermal flux must be minimized.
- **Thermalization depth:** the depth at which neutrons slow to thermal energies in tissue, ideally corresponding to tumor location (typically 2–5 cm).

Through the careful design of the BSA, accelerator-based BNCT systems can achieve clinically effective neutron spectra comparable to reactor-based sources, enabling safe and efficient in-hospital treatment.

4.1.3 Dosimetry

In Boron Neutron Capture Therapy (BNCT), the therapeutic effect arises primarily from the interaction of thermal neutrons with boron-10 atoms that are selectively accumulated within tumor cells. When a thermal neutron is captured by a ^{10}B nucleus, the nuclear reaction $^{10}\mathrm{B}(n,\alpha)^{7}\mathrm{Li}$ takes place, producing an alpha particle and a lithium-7 nucleus. These reaction products are characterized by high linear energy transfer and extremely short ranges in biological tissue, typically comparable to the dimensions of a single cell. As a result, the energy deposition is highly localized, leading to effective tumor cell killing while minimizing damage to surrounding healthy tissue. In the majority of capture events, lithium-7 is produced in an excited state and subsequently emits a prompt gamma photon with an energy of 0.478 MeV, which contributes marginally to the overall absorbed dose.

In addition to the boron capture reaction, neutrons undergo various interactions with the elements that constitute the phantom or biological tissue, giving rise to background dose components. Elastic scattering of neutrons with hydrogen nuclei is particularly significant due to the high hydrogen content of tissue, resulting in recoil protons that

contribute to the neutron dose. Neutron capture by nitrogen-14 through the $^{14}\mathrm{N}(n,p)^{14}\mathrm{C}$ reaction also produces energetic protons, adding to the absorbed dose despite the relatively small cross section of this reaction. Further interactions with oxygen, carbon, and other elements occur through elastic scattering and radiative capture, contributing additional, though smaller, dose components.

Background gamma radiation in BNCT originates both from the incident neutron beam and from neutron capture reactions within the phantom. The dominant gamma contribution arises from neutron capture by hydrogen via the $^{1}\mathrm{H}(n,\gamma)^{2}\mathrm{H}$ reaction, which produces a characteristic gamma photon with an energy of 2.223 MeV. This gamma line represents the principal source of photon background in tissue-equivalent materials. Additional gamma rays are generated by the boron capture reaction at 0.478 MeV and, to a lesser extent, by neutron capture on nitrogen and oxygen, which can produce higher-energy gamma photons in the range of several megaelectronvolts.

Because some of these background gamma photons possess energies exceeding the pair production threshold of 1.022 MeV, electron–positron pair production is physically possible within the phantom. This process is primarily associated with the 2.223 MeV hydrogen capture gamma and the higher-energy capture gammas from nitrogen and oxygen. However, in low atomic number materials such as water or tissue-equivalent phantoms, the probability of pair production remains small compared to Compton scattering. Consequently, Compton interactions dominate photon energy deposition, while pair production contributes only a minor fraction to the total dose. When pair production does occur, the generated positron ultimately annihilates, producing two 511 keV gamma photons that add a negligible secondary contribution to the photon background.

Overall, the absorbed dose in a BNCT phantom results from the combined contributions of the high-LET boron dose, the neutron dose from scattering and nitrogen reactions, and the photon dose arising from capture gamma rays and their subsequent interactions. Accurate modeling of these processes is essential for reliable dosimetry and treatment planning, as the clinical effectiveness of BNCT depends on maximizing the boron-related dose while minimizing background radiation components.

4.2 SPECT tomography for BNCT

4.2.1 System Architecture and Detector used

Different research studies have proposed various detector technologies, including CdTe semiconductor detectors, CZT detectors, and scintillator detectors, to support the clinical adoption of BNCT-SPECT. Conventional SPECT systems use semiconductor or scintillator detectors coupled to specialized mechanical collimators to reconstruct three-dimensional images of the boron concentration within the Region Of Interest (ROI).
A gamma ray deposits energy in the material when it interacts with the detector. This energy excites electrons from the valence band to the conduction band, generating electron-hole pairs. The energy of the incident gamma-ray determines how many electron-hole pairs are created. Detectors measure the electrical signals produced by these electron-hole pairs to compute the gamma-ray's energy. The energy of the incident photons is directly converted into electrical current by these detectors, allowing for accurate photon energy measurement. Prompt gamma rays from both boron reactions at 478 keV and proton-neutron reactions at 2.22 MeV, and the annihilation gamma rays at 511 keV can be measured by using commercially available semiconductor detectors. These detectors are characterized by high energy resolution. Therefore, the two peaks at 478 keV and 511 keV are well separated.
The use of scintillator crystals for a BNCT-SPECT system has been the subject of numerous studies recently due to its outstanding energy resolution at room temperature, detector sensitivity, spatial resolution. The proposed system is visible in Fig. 4.1, a Ce^{3+}-Sr^{2+} co-doped $LaBr_3$ scintillator crystal, coupled with a matrix of 8×8 SiPMs [112, 113]. When a photon interacts with the scintillator crystal, it deposits all or just part of its energy, which is reemitted by the crystal in the form of a pulse of photons. These photons are in turn converted by the SiPMs into an electrical signal, that is then filtered by a gated integrator and converted into an output voltage by the ASICs. The instrument can be connected to a host computer and controlled via a custom designed graphical user interface (GUI). The GUI allows both ASIC setting optimization and data acquisition, which are used to show on the PC the acquired spectrum in real-time.

4.2.2 Tomographic setup

The set-up is developed to perform the SPECT imaging in a BNCT facility. The set-up used for the tomography consists of two supports: one with fixed elements in which boron

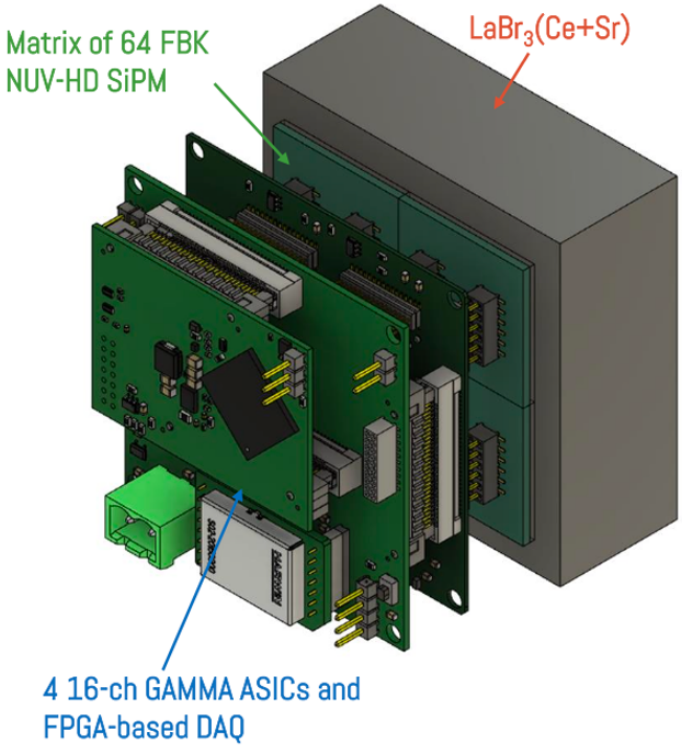


Figure 4.1: Layout of the scintillator[114].

samples are placed in the neutron flux. The second support is a moving part that holds both the detector and collimator aligned with the samples. The collimator is shown in Fig. 4.2.

The distance from the source to the collimator and the distance from the collimator to the detector are 15 cm. The sample are two cylindrical vials filled with boron acid with a concentration of 7371 ppm of ^{10}B, with a inner diameter of 6 mm, 2.6 cm height and a volume of 0.75 ml. The vials are mounted on the opposite vertices of a 1 cm square support. The full configuration is positioned on an aluminium board specifically shaped to fit the set-up. A schematic view is reported in Fig. 4.3. Due to the geometrical constraints limited by the small irradiation room (see Fig. 4.9), the tomographic acquisition is performed in 4 positions. The projection images are aquired only at 0, 60, 120, 180 degree. The images are acquired at two different vials contents in order to pick up the signal coming from boron. The number of cps are estimated directly from spectra under the 478 keV peak. First the two vials are filled with the boron compound at this concentration of 7371 ppm. Next step is to fill vials with same amount of water. The number of cps are estimated directly from spectra under the 478 keV peak, subtracting events of the

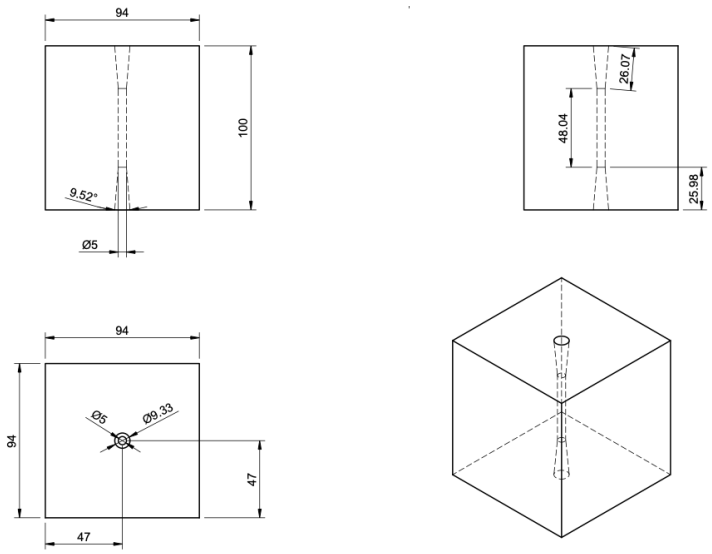


Figure 4.2: Layout of the collimator.

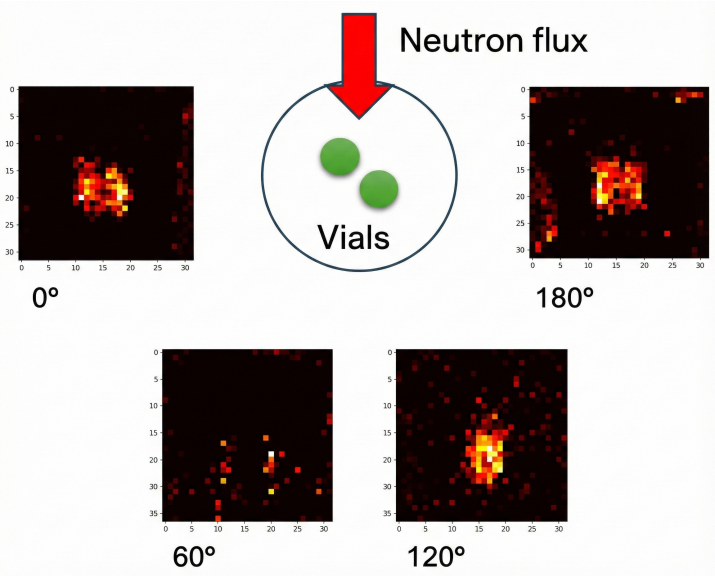


Figure 4.3: Layout of the set-up.

measure with water only from the boron sample measure. Fig. 4.4 shows the spectra of the detector for two different experimental vial contents.

4.3 Database

The iterative models used to perform tomographic reconstruction of the image, could take for the task about 10 minutes. It is crucial in the on-line reconstruction methods to

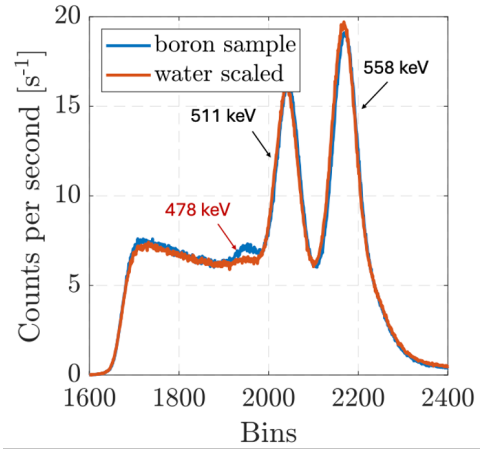
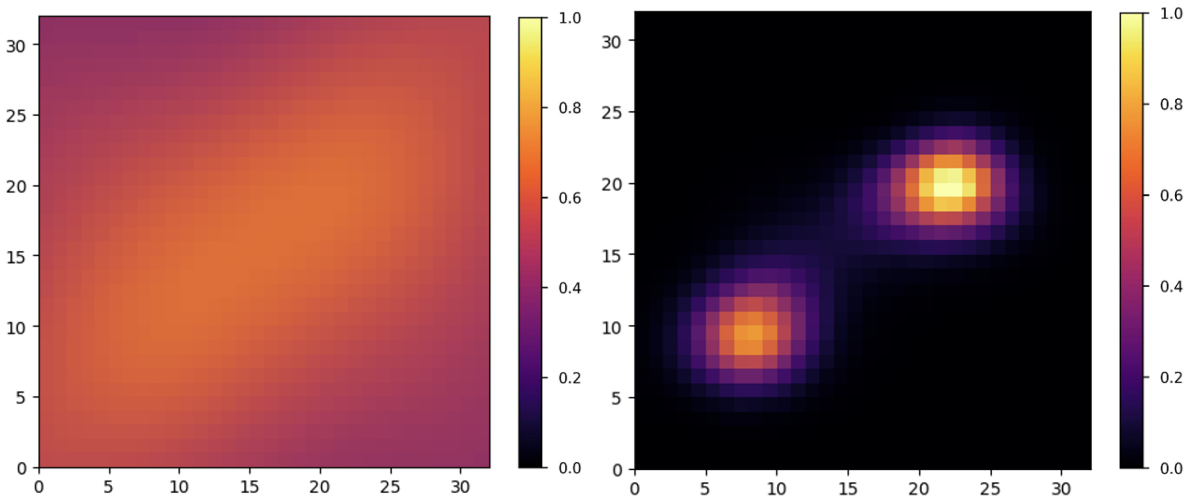


Figure 4.4: Spectra coming from acquisition into two different experimental situations: the red curve is the spectra acquired with boron, while the blue curve is the spectra acquired without boron. The red curve is scaled due to the amount of hydrogen neutron capture reactions that occurs in the background events. The peak at 558 keV is due to the Cadmium foil added to subtract neutrons from camera sides, opening only the acquisition window. The red spectra shows a tiny amount of boron capture reactions, probably a consequence of the thermal neutron interaction with the electronics inside the detector and the possible presence of boron impurities inside the irradiation room [114].

perform the reconstruction istantly, after the acquisition procedure of the images. Most of reconstruction process make use of DL to be performed in an efficient and fast way, but it needs fundamentally two informations: the database from which model learns the underlined principles of the reconstruction, and the model itself.

Because there are very few number of real images, the entire set-up for the tomography is fully simulated in FLUKA [115], taking into account all the mechanical structures that are suited for the acquisition process and the walls of the Laboratory for Nuclear Energy Applications (LENA, Pavia) beam room. The events generated in the scintillator crystal are used to build the projection image, selecting the interval of energy used for the reconstruction of 478 keV gamma (450 keV - 500 keV). The projection stack is a dicom file that is used to perform the reconstruction with OSEM.

The simulation is performed in order to reproduce several sources of 478 keV photons, in different geometrical patterns, in order to train the model in such way it will learn geometrical and energy scale features. Once the reconstructions are obtained, these images could be used as *target* data for the model. The projections used to reconstruct the target of the model, are used to make the first iteration of the algorithm, the so-called *backprojection*. In Fig. 4.5 is shown an example of backprojection and a reconstruction.



(a) (b)

Figure 4.5: (a) Example of backprojection image, the first iteration; (b) The iterated image result by OSEM.

.

This is the starting data with which train model to obtain the reconstruction from the backprojection.

4.3.1 Database from simulation

Monte Carlo simulations were carried out using the FLUKA 2024.1 particle transport code to accurately reproduce the Prompt Gamma Neutron Activation Analysis (PGNAA) beamline at LENA and compare simulated detector responses with experimental measurements. The implemented model includes a full representation of the irradiation room, shielding, collimation system, detector, and test phantoms, allowing a realistic simulation of neutron and gamma particle transport under the specific conditions of the LENA facility. The first simulation is performed with another simulation software, called Geant4 [116], that is particularly used for simulate the geometries and materials of detectors used in the high energy physics laboratory, like the work described in Appendix A, where a facility at *Conseil européen pour la recherche nucléaire* (CERN) is fully simulated in Geant4.

The Geant4 simulation performed in the thesis focuses on a stationary SPECT configuration designed to evaluate detection efficiency for specific capture or emission scenarios. The modeled geometry consists of four independent scintillator detector modules arranged in a fixed array. These modules are positioned at 90° intervals relative to one

another, effectively creating a square detection geometry that surrounds the field of view. Placed precisely at the geometric center of this four-detector arrangement is a phantom containing boron. The simulation analyzes the radiation transport and signal acquisition from this boron-doped phantom, calculating the system matrix and sensitivity profile provided by the four orthogonal, static scintillator heads. Fig. 4.6 shown an illustration of the set-up. Using this set-up, are tested two different boron distribution: 5 spheres

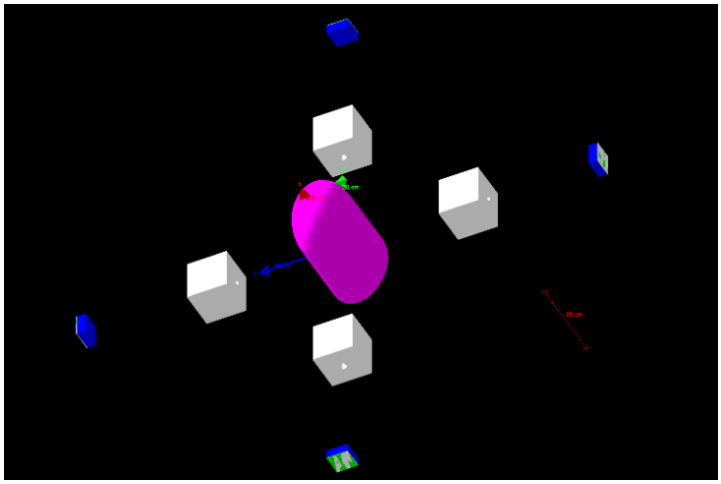
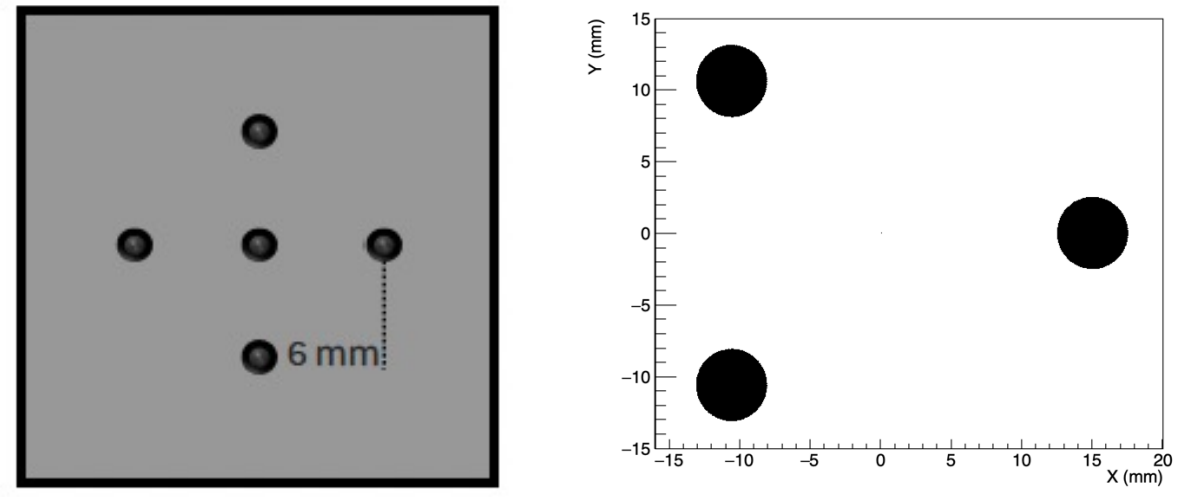


Figure 4.6: Detection system implemented in Geant4, including scintillator, collimator and hypothetical phantom.

(source 1) and 3 spheres (source 2). The source's profiles are shown in Fig. 4.7. The hits



(a) (b)

Figure 4.7: Test sources: (a) the Source 1, made by 5 spheres; (b) the Source 2, made by 3 spheres.

of the gamma in the detector of the Geant4 simulations, are shown in Fig. 4.13 and 4.14, where the two sources are detected with the detector that turns around every 22,5°.

The second set of simulations for the thesis is performed with Fluka software. The simulations are performed to really simulate the boron vials experiments, in which the boron compound is mixed with water, with a concentration of 7371 ppm. Neutron transport was performed using FLUKA's pointwise treatment based on ENDF nuclear data libraries, with a minimum neutron energy threshold of 10^{-14} GeV. The production and transport of secondary particles including photons, electrons, recoil protons, and heavy nuclei produced in the $^{10}\mathrm{B}(n,\alpha)$ reaction were subject to a lower energy cut of 10 keV.

The initial simulations reproduced the experimental configuration of the phantom, collimator, and detector without modification, in Fig. 4.8. The irradiation room is modeled as a parallelepiped with internal dimensions $80 \times 200 \times 220$ cm^3, enclosed by 50 cm-thick reactor wall and barite-concrete shielding, shown in Fig. 4.9. The compositions of these materials are given in Tables 4.1 and 4.2.

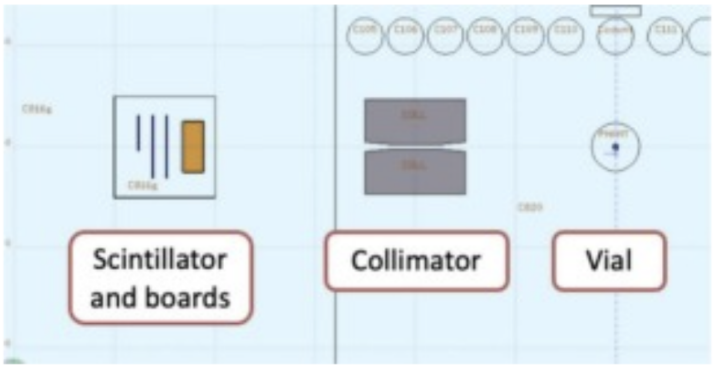


Figure 4.8: Detection system implemented in FLUKA, including scintillator, collimator, and support structures [117].

The detection system consists of the scintillator crystal, its holder, and the associated electronics boards (motherboard and powerboard). The boards, each measuring 9.4×9.4 cm^2 with thicknesses of 1.6 mm and 1.2 mm respectively, were included in the model geometry; individual surface-mounted components were omitted because their interaction probability with thermal neutrons is negligible.

The detection system incorporates a channel-edge pinhole collimator with a 5 mm

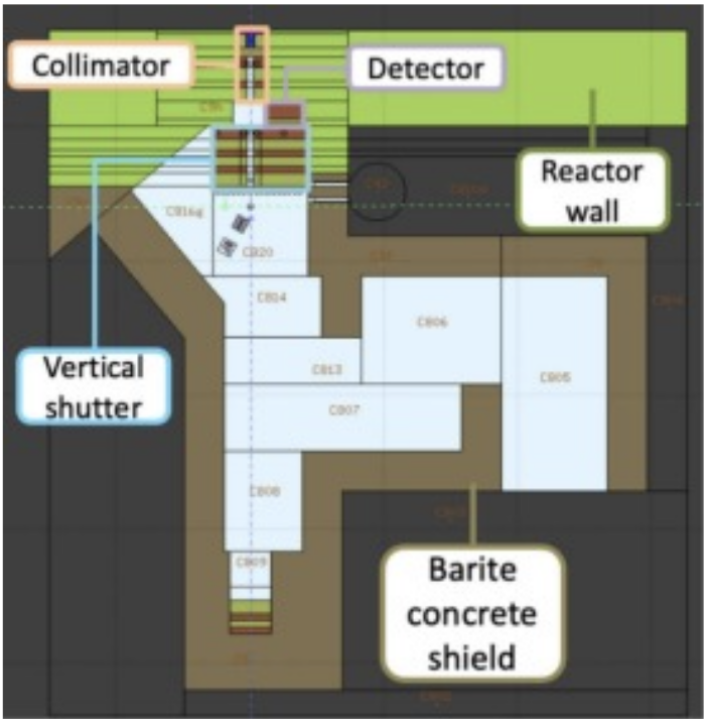


Figure 4.9: Top view of the PGNAA line at LENA, showing irradiation room, shielding, detector, collimator, and phantom.[117]

diameter aperture. At a distance of 30 cm, the collimator provides an acceptance angle of 9.52°, corresponding to a 5 cm field of view. In this work the effective field of view is slightly reduced by geometric constraints.

Because the neutron fluence rate at the LENA PGNAA facility is far lower than in clinical BNCT beams, a boron concentration of 7371 ppm at LENA corresponds approximately to 44 ppm under clinical fluence, making the experimental configuration representative of treatment-relevant concentration ranges.

The vial position is the central position aligned with the collimator aperture. The shielding materials of the PGNAA room include barite concrete and reactor-wall concrete (Tables 4.1–4.2). Their compositions affect neutron attenuation and, consequently, the background gamma spectrum.

A significant potential source of unwanted background is boron present in FR4, the standard substrate used in printed circuit boards. FR4 is a flame-retardant composite consisting of:

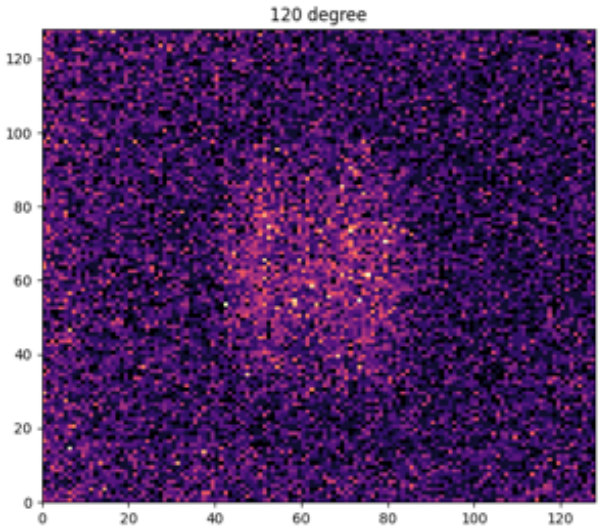


Figure 4.10: The projection at angle equal 120 for the vial source.

- an epoxy resin matrix (composition in Table 4.3),
- woven E-glass fibers (composition in Table 4.4).

The glass fibers contain boron in the form of B_2O_3, typically 5–10% by weight. Although this seems minor, boron has a large capture cross section for thermal neutrons, and the resulting neutron capture reaction produces 478 keV gamma rays. This energy is identical to the signal gamma from the $^{10}B(n,\alpha)^7Li$ reaction, making FR4 a critical background source that must be evaluated and mitigated. FLUKA tracks particles until their total energy is deposited in the scintillator. For each interaction, the spatial coordinates and energy deposition are stored. A strict acceptance window from 465 keV to 490 keV is applied to isolate gamma rays associated with the 478 keV line. In Figure 4.10 there is a single projection at angle equal 120. For image construction, only the y and z coordinates of valid interactions are retained. A schema of this process is shown in Fig. 4.11.

For tomographic imaging, the geometry was modified to acquire multiple angular projections. Two vials, each containing 7371 ppm of boron, were placed at opposite corners of a 1 cm square support, separated by approximately 1 cm. To improve the detection efficiency, the distances between the rotation center, collimator, and detector were reduced to 15 cm.

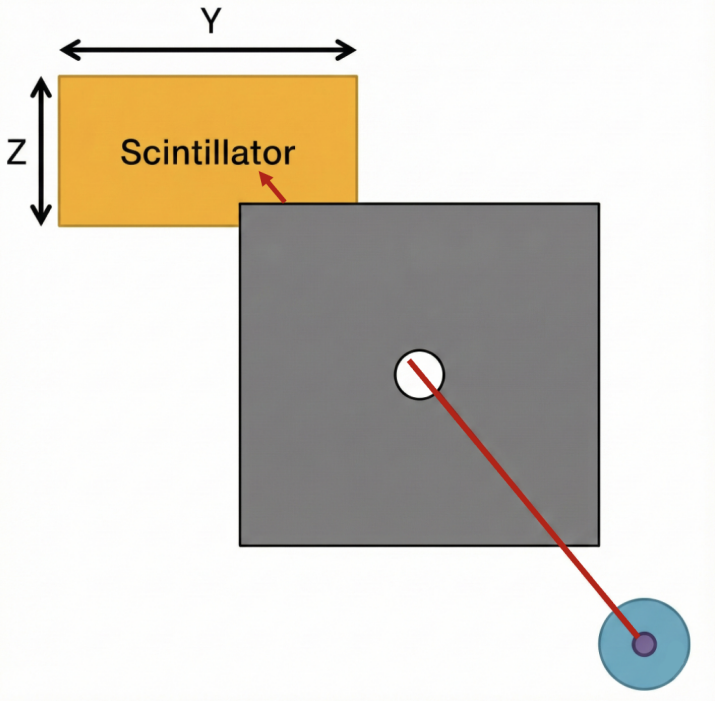


Figure 4.11: Schematic representation of image formation from 478 keV gamma interactions in the scintillator [117].

The detection system was rotated counterclockwise about the vials to acquire projections at four angles: 0°, 60°, 120°, and 180°, as shown in Fig. 4.12. While in Fig. 4.13 are shown the projections of the source 1, and in Fig. 4.14 are shown the projections of the source 2.
The setup simulated for the two vials object reconstruction is resumed in Fig. 4.15.

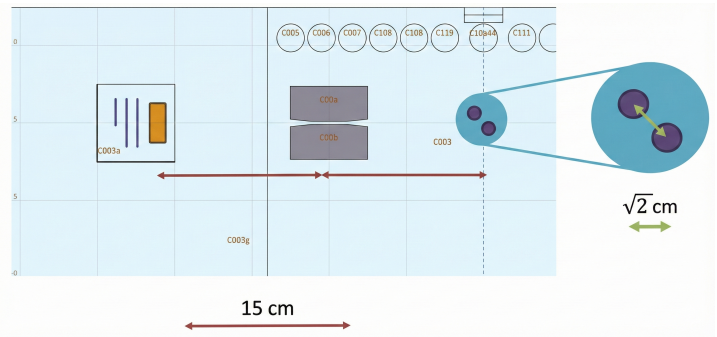


Figure 4.12: Tomographic setup geometry. The detection system rotates counterclockwise around the vials to obtain four projections [117].

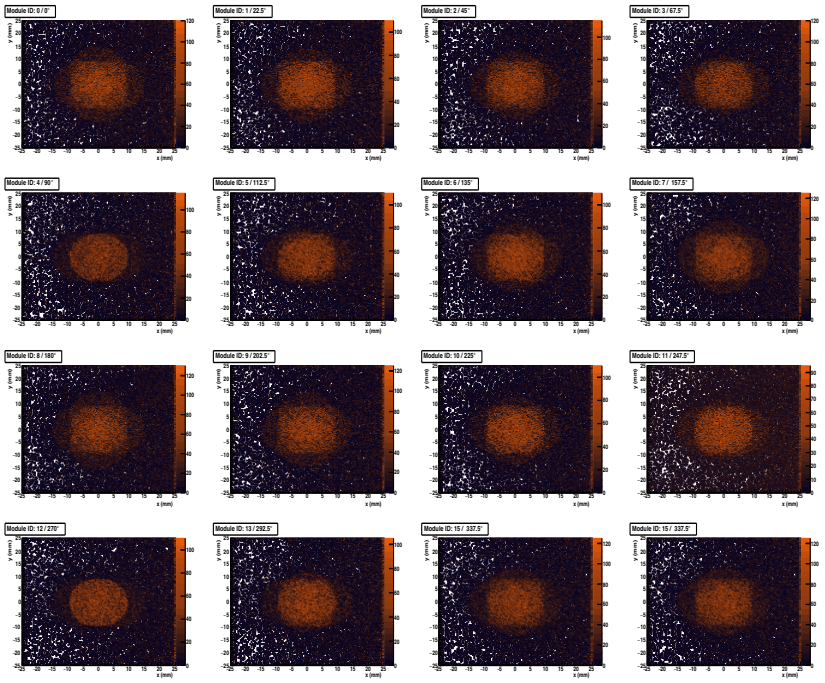


Figure 4.13: Tomographic projections of Source 1. The projections are 16, at angles equally separated in 360°. The one used are only the 4 angles at 90° intervals.

Table 4.1: Composition of barite concrete used in the PGNAA shielding.

Material	**Weight fraction** [%]
Hydrogen	1.7864
Carbon	0.007692
Oxygen	38.83776
Magnesium	0.015385
Aluminium	0.2605539
Nitrogen	2.59247
Sulfur	10.5684
Potassium	0.1
Calcium	0.338462
Iron	0.10769
Barium	45.26216

4.3.2 Database from measurements

This section describes in detail the experimental configuration used to perform BNCT–SPECT tomographic measurements at the TRIGA MARK II reactor at LENA.

The set-up used for the tomography consists of two supports: one with fixed elements in which boron samples are placed in the neutron flux. The second support is a moving

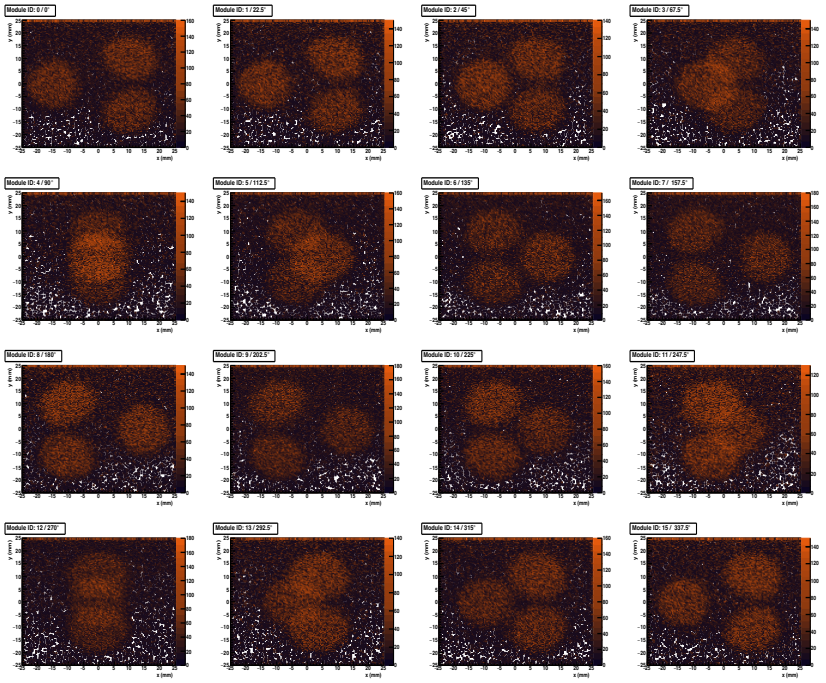


Figure 4.14: Tomographic projections of Source 2. The projections are 16, at angles equally separated in 360°. The one used are only the 4 angles at 90° intervals. Is possible to appreciate better the difference between the spheres, in contrast to Source 1 projections.

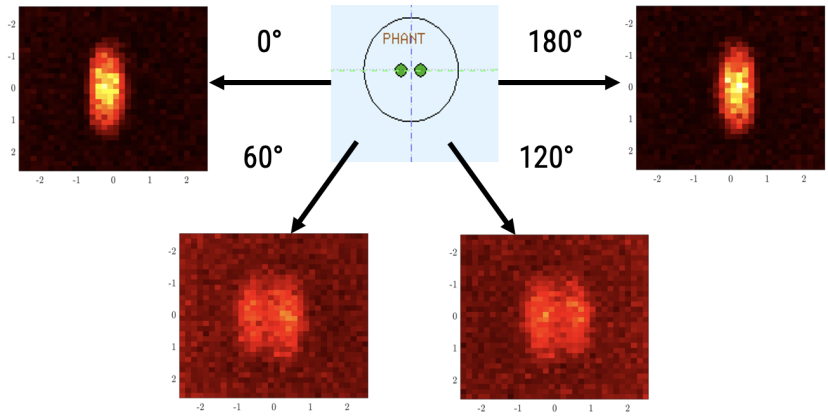


Figure 4.15: Tomographic system for vials simulation. The set-up shows the projection acquired. Is possible to appreciate that the projections in the forward direction to the neutron beam are more noisy then the left and right projections, due to more background.

part that holds both the detector and collimator aligned with the samples. The distance from the source to the collimator and the distance from the collimator to the detector are 15 cm. The sample are two cylindrical vials filled with boron acid with a concentration of 7371 ppm of ^{10}B, with a inner diameter of 6 mm, 2.6 cm height and a volume of 0.75 ml. The vials are mounted on the opposite vertices of a 1 cm square support. The

Table 4.2: Composition of the reactor wall concrete.

Material	**Weight fraction [%]**
Hydrogen	0.56
Oxygen	49.56
Nitrogen	31.35
Aluminium	4.56
Iron	1.22
Magnesium	0.24
Sodium	1.71
Potassium	1.92
Sulfur	0.11
Calcium	8.26

Table 4.3: Composition of epoxy resin used in FR4 (FLUKA default).

Material	**Weight fraction [%]**
Hydrogen	47.5
Carbon	45.0
Oxygen	7.5

Table 4.4: Composition of glass fiber reinforcement in FR4.

Material	**Weight fraction [%]**
SiO_2	55.2
Al_2O_3	14.8
B_2O_3	7.3
MgO	3.3
CaO	18.7
K_2O	0.2
Na_2O	0.2
Fe_2O_3	0.2
F	0.1

full configuration is positioned on an aluminium board specifically shaped to fit the set-up. Due to the geometrical constraints limited by the small irradiation room, the tomographic acquisition is performed in 4 position. The projection images are aquired only at 0, 60, 120, 180.

Figure 4.16 illustrates the essential components of the setup. The mechanical structure of the SPECT can be described starting from the geometric arrangements. The SPECT system consists of two custom mechanical structures fabricated from polylactic acid (PLA):

1. a *fixed support* used to position and align the boron-containing samples with the incident neutron flux, and

2. a *moving support* supporting both collimator and detector, ensuring their coaxial alignment with the samples during rotation.

Due to space constraints in the irradiation room, both the sample–collimator distance and the collimator–detector distance were set to 15 cm, which is smaller than the standard configuration normally used with this detection module. At these distances, the pinhole collimator defines a projected field of view (FoV) of 5.0 cm diameter at both source and detector planes.

The two cylindrical vials served as radioactive targets. The boron concentration is chosen to offset the low neutron flux available at the facility: the maximum achievable flux, 3.14×10^6 n/cm^2/s, is about three orders of magnitude smaller than clinical BNCT flux values ($\sim 10^9$ n/cm^2/s). Consequently, the reaction rate corresponds roughly to that expected from a 20 ppm ^{10}B concentration under clinical irradiation.

The two vials are mounted at opposite vertices of a square PLA support with a side length of 1 cm.

A larger water phantom, sometimes used in BNCT imaging experiments, is not used here because the available neutron beam is predominantly thermal (97% thermal component). Immersing the vials in water would cause strong thermal-neutron absorption and scattering on hydrogen, preventing sufficient neutron flux from reaching the ^{10}B-containing sample.

To suppress activation of the detector front-end electronics and the scintillator crystal by thermal neutrons, the detection module is enclosed in a cadmium shielding case (Fig. 4.16). Cadmium is chosen for its extremely high thermal-neutron capture cross section ($\sigma_{^{113}\mathrm{Cd}} = 19852$ barns). The effectiveness of this shielding configuration is previously validated via simulations and measurements.

Additional lead foils were inserted between the Cd shielding and the lateral surfaces of the detector to attenuate capture gammas emitted from cadmium, in particular the dominant 558 keV and 651 keV lines. Moreover, lead blocks of dimensions 10 cm × 10 cm × 5 cm are installed between the Beam port and the mechanical setup to suppress

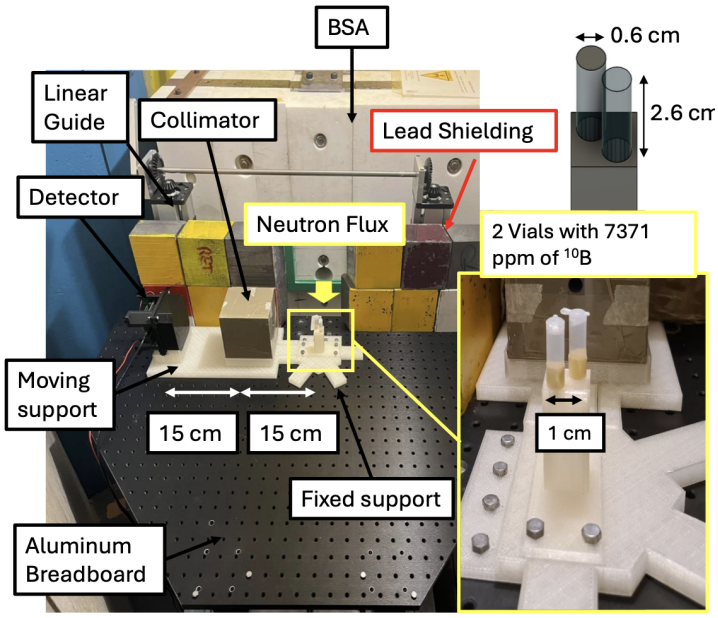


Figure 4.16: Schematic of the BNCT–SPECT setup, including fixed and moving PLA supports, rotation stage, shielding components, and detector wrapped in Cd foils [114].

background gamma radiation originating from interactions of neutrons with the BSA.

The complete setup is mounted on a custom-shaped aluminum optical breadboard that fits the irradiation chamber. The breadboard is attached to an internal linear guide and two height-adjustable steel legs, enabling fine vertical alignment so that the center of the vials coincided with the neutron beam axis.

Spatial limitations inside the irradiation room restricted the experiment to a *partial-angle* tomographic acquisition. For this reason, projection images are acquired at only four angles:

$$0^\circ,\ 60^\circ,\ 120^\circ,\ 180^\circ,$$

as illustrated schematically in Figure 4.17.

At each of the four angular positions, two types of measurements were taken: Signal measure, that consists of vials filled with the 7371 ppm ^{10}B solution, and the Background, that are the identical vials filled with distilled water.

The experimental campaign is conducted with a nuclear power of 70 kW : corresponding to 0.87×10^6 n/cm^2/s, with an acquisition time of 36 min per angle. For each projection, 2-D images were reconstructed and the number of 478 keV γ-ray events

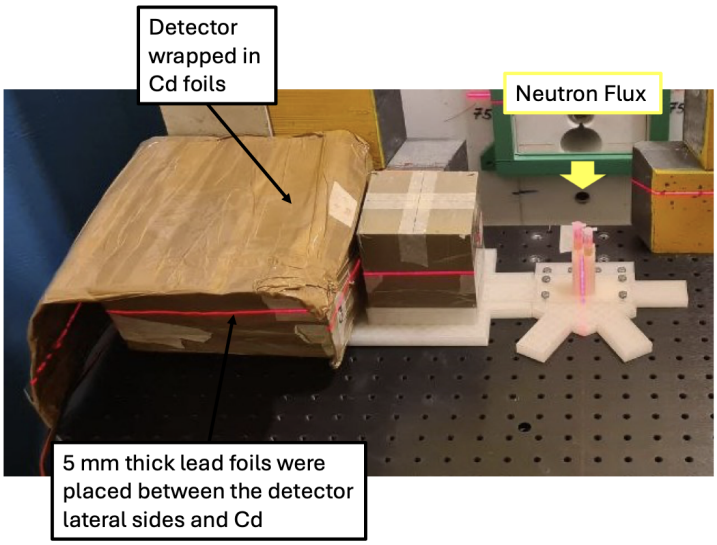


Figure 4.17: Acquisition geometry showing the setup of the measurements with the cadium foil. The right alignment of the tomography setup is performed using a level laser method [114].

(from the $^{10}\mathrm{B}(n,\alpha)\gamma$ reaction) are extracted. The detected $478\,keV$ count rate (cps) is computed from the spectra by subtracting the background spectrum from the boron-sample spectrum and integrating under the $478\,keV$ peak. A fitting model composed of three Gaussian functions (representing the 478, 511, and 558 keV peaks) plus a linear background is used.

All events are processed through the ANN-based position estimator. Only events whose reconstructed interaction coordinates fell inside the collimator FoV (a $12.5\,mm$ radius circle at the detector center) are retained.

The experimental results are then compared with Monte Carlo simulations performed using the FLUKA code.

Each image is reconstructed by mapping the ANN-estimated (x, y) coordinates of events within the $478\,keV$ window ($\pm 17\,keV$) onto a 32×32 pixel grid spanning the $5\,cm$ detector surface (pixel pitch: $1.56\,mm$).

An unexpected effect is observed: the distilled-water background spectra yielded a higher measured Output Count Rate (OCR) than the ^{10}B-containing measurements, consistent with Monte Carlo predictions. This is likely due to the fact that neutrons are partially absorbed by the boron solution, reducing room-scattered background.

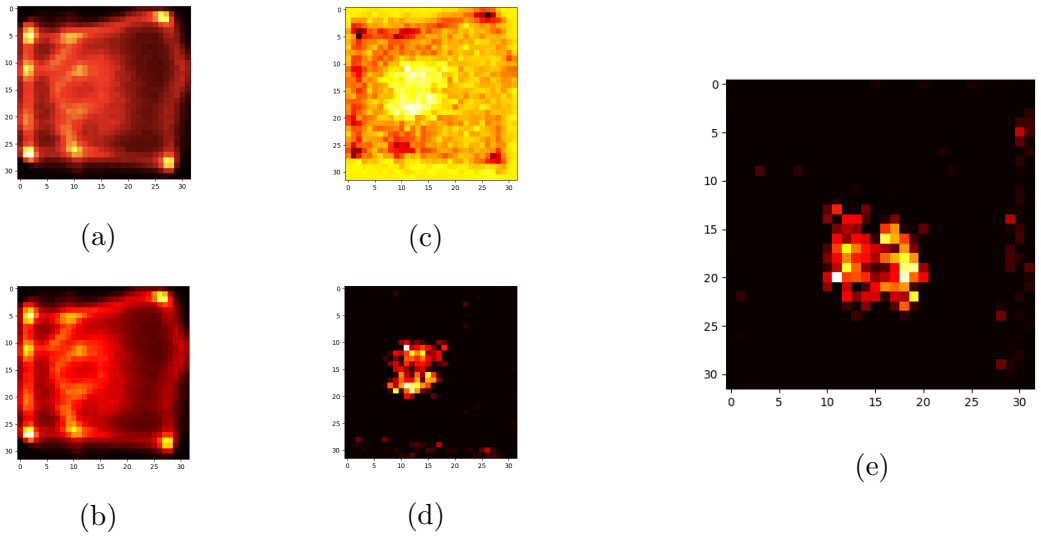


Figure 4.18: The pre-processing consists of taking the difference between the image with boron filled vials (a) and the image without boron (b), subtract the two images (c) and substitute with zeros the negative values (d), finally rotation is to get the correct orientation (e).

To correct this, a normalization factor is defined. First, the Input Count Rate (ICR) is computed from the OCR using the detector dead time DT = 7.4 μs:

$$\mathrm{ICR} = \frac{\mathrm{OCR}}{1 - (\mathrm{OCR} \cdot \mathrm{DT})} \tag{4.5}$$

The correction factor is

$$\mathrm{CF} = \frac{\mathrm{ICR}_{\mathrm{Boron}}}{\mathrm{ICR}_{\mathrm{Water}}} \tag{4.6}$$

The background image is multiplied by CF before pixel-wise subtraction.

The spatial resolution is estimated by summing pixels along columns and fitting the resulting 1-D profile with two Gaussians. The expected resolution (8.4 mm) is computed analytically by modeling each vial as the sum of five Gaussian-distributed sources (1-mm spacing), matching the projection geometry described in [118]. The number of cps are estimated directly from spectra under the 478 keV peak. First the two vials are filled with the boron compound at this concentration of 7371 ppm. Next step is to fill vials with same amount of water. The number of cps are estimated directly from spectra under the 478 keV peak, subtracting events of the measure with water from the boron sample measure. In Fig. 4.18 is possible to appreciate the difference between the acquired vials with and without boron and with background subtraction.

Four background corrected projections (at the four acquisition angles) are fed into the OSEM reconstruction in the Pytomography framework. The reconstruction uses MLEM with a number of iterations chosen empirically to avoid divergence due to the limited angular sampling. A Point Spread Function (PSF) correction model is applied to jointly invert the geometric PSF of the pinhole collimator and the intrinsic PSF of the detector.

From the 3-D reconstructed volume, top and lateral views are obtained by summing along the z-axis and x-axis respectively. The vial separation is quantified by fitting Gaussian profiles along both axes and extracting the centroid positions and FWHM values.

4.4 Tomography by iterative approach

The modern frameworks are based on classical approaches. Advanced tomographic frameworks further extend iterative reconstruction by incorporating regularization terms or prior information to stabilize the solution and suppress noise amplification. These approaches include penalized likelihood methods and Bayesian reconstructions that use anatomical priors from complementary modalities such as CT or MRI, as well as edge-preserving or sparsity-promoting constraints. In hybrid SPECT/CT systems, attenuation maps derived from CT are routinely integrated into the reconstruction process, reinforcing the physical accuracy of the tomographic model. Overall, contemporary SPECT image reconstruction relies on statistically grounded tomographic frameworks that tightly couple physics based system modeling with iterative optimization, enabling clinically robust and quantitatively reliable imaging. `PyTomography` framework provides a highly modular and well-structured environment for conducting iterative SPECT reconstruction. Its core elements accurate system modeling, explicit handling of collimator resolution, robust forward/adjoint operators, flexible likelihood definitions, and DICOM interoperability enable both research-oriented and clinical-grade algorithm development. Through its Expectation maximization-based solvers and support for advanced penalization and PSF modeling, the package allows construction of physically accurate SPECT reconstruction pipelines optimized for resolution, noise, and quantitative accuracy.

This section presents a detailed overview of an iterative reconstruction approach for SPECT based on the `PyTomography` Python framework. The package provides a modular interface for system modeling, tomographic projection, likelihood-based objective functions, detector–collimator response modeling, and algorithmic tools for MLEM,

OSEM, and penalized variants for improved regularity.

4.4.1 System Model and Forward Projection

At the core of the reconstruction process lies the *system matrix* A, which maps the unknown radiotracer distribution $x(\mathbf{r})$ to the predicted projection measurements $y(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ denotes the SPECT acquisition geometry (angle, detector coordinates, and energy window). In `PyTomography`, this mapping is implemented by a flexible forward projector of the form

$$y = Ax + b \tag{4.7}$$

where b represents scatter, background counts, or modelled additive noise. The forward operator handles geometric projection under pinhole or parallel-beam SPECT, collimator point-spread function (PSF), can introduces in the model the detection efficiency, can simulate a solid-angle weighting and also an energy-window acceptance. The forward operator is typically implemented as a Python class that encapsulates both the projection geometry and the detector–collimator response. Using the software's vectorized operations, the user provides: a 3-D image volume populated by voxelized activity, a description of angular sampling and camera radius, the collimator model (e.g. Gaussian-blur approximation, geometric-response kernels), optional attenuation maps or CT-derived μ-maps. The result is a synthetic sinogram that matches the format and dimensionality of the acquired SPECT projection data. Because the package is GPU-compatible, forward projection can be carried out efficiently for large or multi-angle data sets.

4.4.2 Backprojection and Adjoint Operator

The adjoint operator A^* plays a central role in iterative reconstruction. In the context of SPECT, the backprojection distributes projection-domain residuals (ratio of measured to predicted counts) back into image space. Formally, the adjoint maps a projection-space vector $r(\boldsymbol{\theta})$ to a volumetric correction term:

$$A^* r = \text{backproject}\big(r(\boldsymbol{\theta})\big) \tag{4.8}$$

In `PyTomography`, the backward operator performs: angular backprojection consistent with acquisition geometry, PSF-adjoint convolution to reverse collimator blurring, normalization by the sensitivity image, optional attenuation compensation via exponential weighting. The framework ensures that A^* is implemented as the true mathematical

adjoint of A, an essential requirement for convergence of Expectation Maximization algorithms.

4.4.3 DICOM Utilities and Data Handling

To interface with clinical workflow and SPECT/CT acquisition systems, `PyTomography` includes utility routines for reading, writing, and managing DICOM projection data. These tools allow to perform the parsing of DICOM SPECT projection stacks (transmission or emission), the extraction of metadata such as energy windows, angle tables, detector radius, pixel spacing, and collimator model, the possibility to add an integration of CT-derived attenuation maps to compute linear attenuation coefficients $\mu(\mathbf{r})$, and the feature for exporting reconstructed volumes to DICOM or NIfTI format, and finally performing energy-window filtering and preprocessing. The DICOM utilities abstract modality-specific formatting details, enabling a unified workflow regardless of vendor (GE, Philips, Siemens). Consistency checks ensure that angles, pixel dimensions, and frame ordering are matched between projection data and the forward operator.
The projections are pre-processed in a dicom file, which is fed into OSEM algorithm performed with the use of Pytomography library. In the dicom were set different parameter in the metadata such as number of projection, distance from the center of image, pixel size, FOV, ecc.

4.4.4 Poisson Likelihood Model

The SPECT photon-counting follows Poisson statistics. That is the reason why the likelihood function for the measured data y given the predicted projections Ax is

$$\mathcal{L}(x) = \prod_i \frac{(Ax+b)_i^{y_i}}{y_i!} \exp\bigl(-(Ax+b)_i\bigr) \tag{4.9}$$

where the index i spans detector pixels and projection angles. Maximizing the log-likelihood is equivalent to minimizing the Kullback–Leibler divergence between measured and estimated sinograms:

$$\Phi(x) = \sum_i \left[y_i \ln \frac{y_i}{(Ax+b)_i} + (Ax+b)_i - y_i \right] \tag{4.10}$$

The Expectation Maximization update rule derived from this likelihood is:

$$x^{(k+1)} = x^{(k)} \frac{A^* \left(\dfrac{y}{Ax^{(k)}+b} \right)}{A^*\mathbf{1}} \tag{4.11}$$

While for the OSEM, that accelerates convergence by splitting the projection set into subsets S_j, update rule derived from this likelihood is, at each iteration:

$$x^{(k+1)} = x^{(k)} \frac{A^*_{S_j}\left(\frac{y_{S_j}}{A_{S_j}x^{(k)}+b}\right)}{A^*_{S_j}\mathbf{1}} \tag{4.12}$$

`PyTomography` supports arbitrary subset partitioning and automatically manages angle-grouping. All components of this update—forward projection, residual formation, and backprojection are explicitly implemented in the package, enabling both MLEM and OSEM strategies.

4.4.5 Resolution and Collimator Response Modeling

A distinguishing feature of `PyTomography` is explicit system modeling of the collimator resolution. The resolution function typically incorporates the geometric response of the collimator holes, the septal penetration, the distance-dependent blur and the detector intrinsic resolution. For pinhole SPECT, the PSF may be approximated by a depth-dependent Gaussian or tabulated analytically via collimator geometry. For parallel or fan-beam collimators, the distance-dependent resolution is

$$R_{coll} = \sqrt{R_0^2 + \left(\frac{d}{f}\right)^2} \tag{4.13}$$

with R_0 being intrinsic detector resolution and f the geometrical focal length. While for pin-hole collimator the distance-dependent resolution is

$$R_{coll} = \left(d + \frac{ln(2)}{\mu}\tan(\frac{\alpha}{2})\right)\frac{(l+b)}{l} \tag{4.14}$$

with d being the hole diameter, μ the attenuation coefficient for the gamma attenuation in lead material, which have to be set to the value at 478 keV, that is 1.31 cm^{-1}, α the angle aperture of the hole, which in the pin-hole used is $\alpha = 9.52$.

The PSF convolution is incorporated directly into the forward operator:

$$(Ax)(\boldsymbol{\theta}) = \left(h_{\text{PSF}} * \text{project}(x)\right)(\boldsymbol{\theta}) \tag{4.15}$$

and the adjoint operator applies the transpose convolution as part of backprojection. More advanced collimator responses (e.g., Monte Carlo kernels, depth-dependent anisotropic PSFs, Gaussian-weighted Modified Uniformly Redundant Array apertures) may also be

used, as the framework supports arbitrary convolution kernels.
A further study on the reconstruction, shows that the Point Spread Function (PSF) depends mostly from the collimator channel, and not from angle aperture. The new function used is

$$R_{coll} = (\frac{x}{f})\sqrt{R_0^2 + (\frac{f+x}{x+\frac{CH}{2}}) * d^2} \tag{4.16}$$

where $R_0 = 3.0$ mm is the intrinsic resolution, $f = 30$ cm is the distance collimator-detector, $x = 30$ cm is the distance collimator-source, $d_e = 5$ mm is the pinhole diameter, $CH = 48.04$ mm is the channel length. The two different version of the formula shows different results.

4.4.6 Results with Geant4 simulation

In Fig. 4.19 the results are shown for the test source 1 with the formula 4.14. In Fig. 4.20 the results are shown for the test source 2 with the formula 4.14. In Fig. 4.21 the results are shown for the test source 2 with the formula 4.16.
Comparisons between point-like sources and spherical distributions revealed distinct variations in the photon acceptance angles for each detector module. While the 90° symmetric arrangement offered uniform sensitivity for centrally located spherical sources, the results indicate significant anisotropy in the count rates for elongated or off-center irregular geometries. Consequently, the static nature of the four-detector array introduces specific artifacts depending on the source's orientation, suggesting that accurate reconstruction for complex boron geometries requires precise calibration of the system matrix to account for these geometric non-uniformities. The PSF formula used in Fig. 4.20 is an approximation of the formula used in Fig. 4.21, but that approximation happens when the channel lenght is small compared to the distance collimator-detector. That results shown reflect that approximation: the non-perfect circles in Fig. 4.20 are due to the abscence of the rays coming from the outer zone of the FOV.

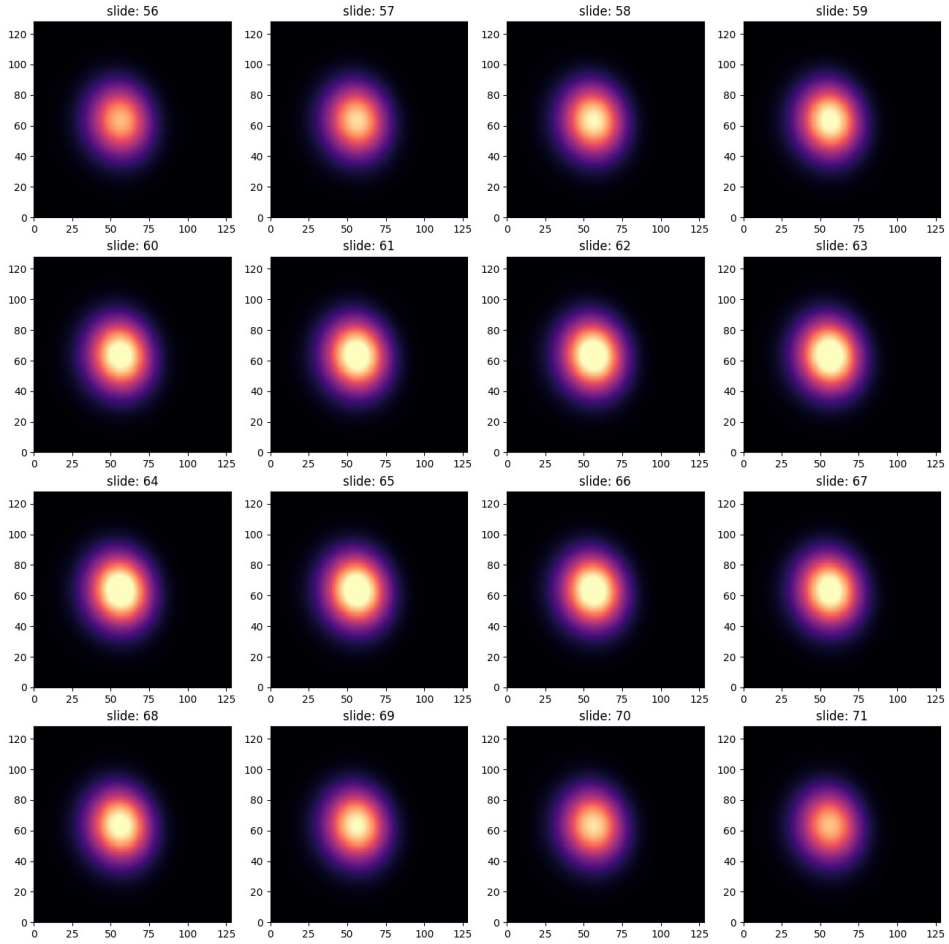


Figure 4.19: Results with the fixed set-up and Source 1 geometry. This results shows that the resolution is not sufficient to perform the Source 1 reconstruction, because the spheres are too small and the distance is small.

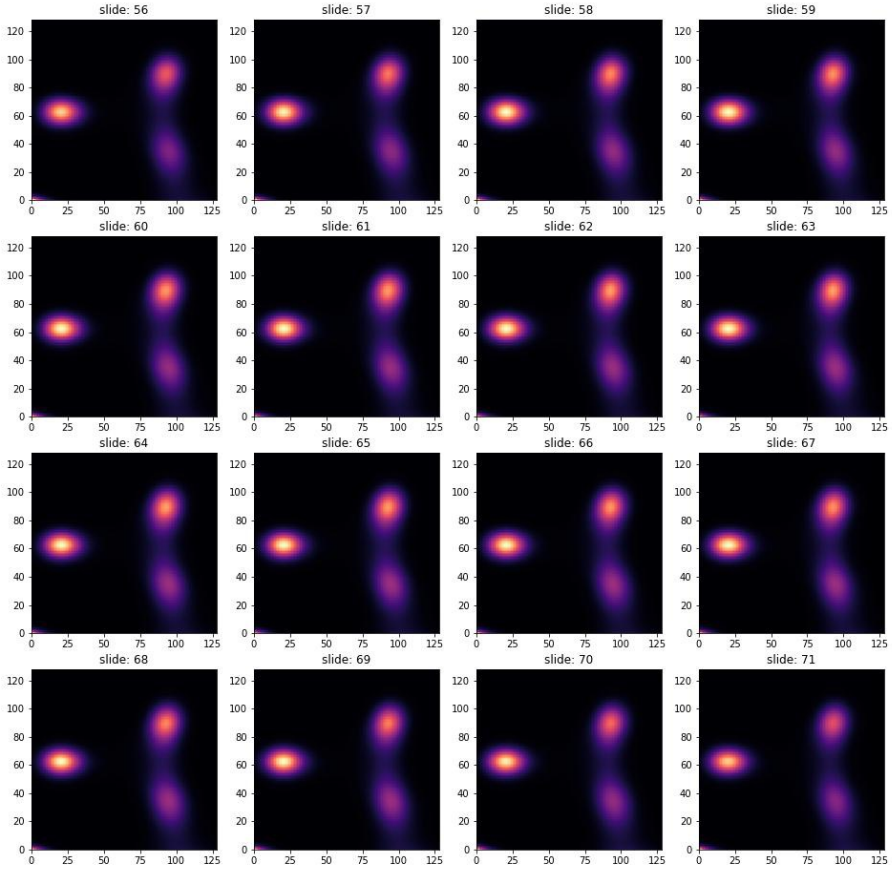


Figure 4.20: Results with the fixed set-up and Source 2 geometry. The results shows a good agreement with Source 2. The function of the resolution used is the responsible of the blurred shapes and not true dimensions.

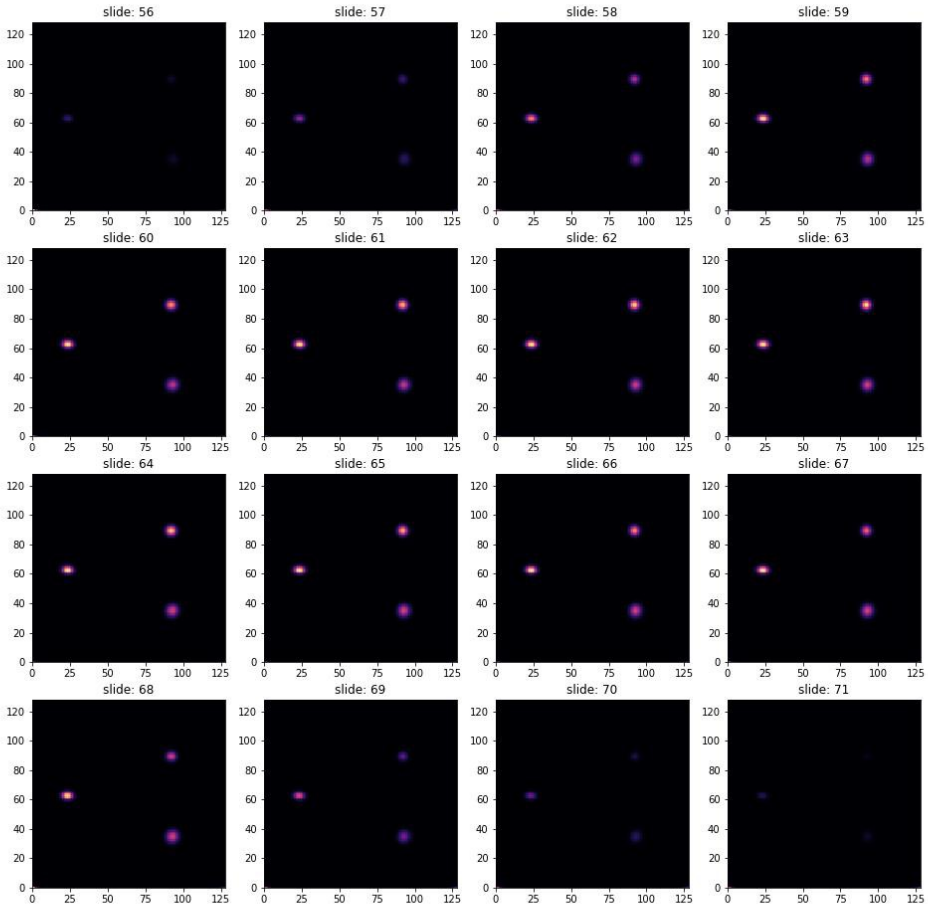
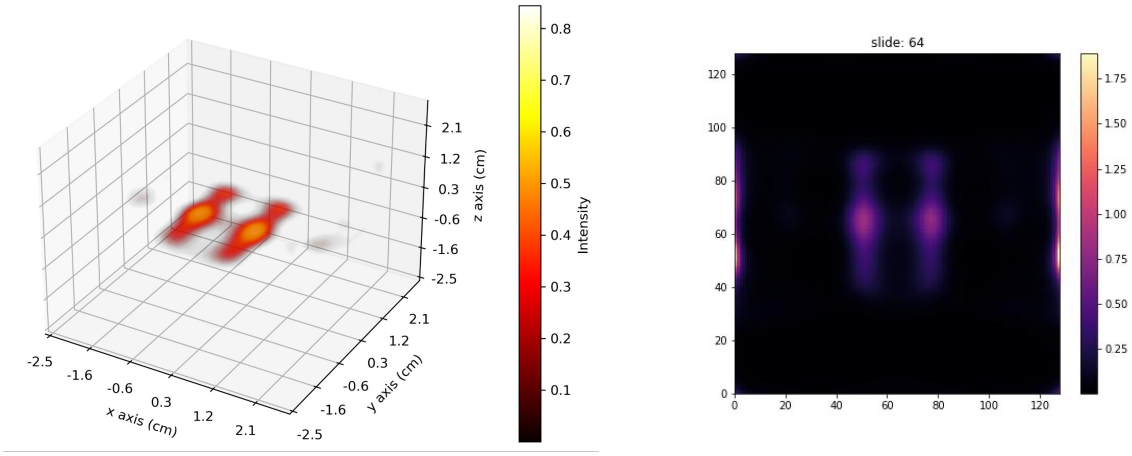


Figure 4.21: Results with the fixed set-up and Source 2 geometry, but with updated formula. The new formula shows very good agreement with Source 2, the spheres are less blurred.

4.4.7 Results with Vials object

The tomographic reconstruction of the simulated vials are shown in the following results. In order to obtain them, the OSEM used 50 iterations, with subset 3, and the resolution formula 4.16. The result is shown in Fig. 4.22.



(a) (b)

Figure 4.22: The result of the reconstructed simulated vials with OSEM at 50 iterations, subset 3: (a) a 3-dimension plot, (b) a slice in the xy plane center on the vials.

The following results are the reconstruction obtained with the measurements. In the experimental campaign the vials are filled with a certain amount of boron compound and the expected result is the reconstruction of the spatial distribution of the boron in the vials. In Fig. 4.23 is shown the reconstruction tomography of the vials with no corrections factor calculated in 4.6. While the Fig. 4.24 shows results of tomography with correction factors applied to the projections, and the OSEM iterated 20 times. In Fig. 4.25 is shown the result of tomography with correction factors and 50 iterations. The results highlight a critical consequence of the adopted reconstruction technique: the reliability of the spatial distribution of the boron concentration is intrinsically linked to the accuracy of the background subtraction procedure. This dependency implies that the reconstruction does not solely reflect the true boron signal, but rather the differential response between two measurement conditions. Consequently, the experimental protocol must be strictly structured as a two-step acquisition process. A reference measurement without the boron compound is not optional, but essential, as it defines the baseline upon which the boron-related signal is extracted. Any deviation, instability, or inconsistency in this reference acquisition directly propagates into the reconstructed boron map. Furthermore, the analysis demonstrates that the background correction factor does not act as a simple

global scaling parameter. Instead, it influences the reconstructed image at the pixel level, thereby affecting both quantitative values and spatial contrasts. This indicates that correction factors are coupled with local image features and noise characteristics, and their impact cannot be assumed to be spatially uniform. As a result, an improper estimation of the background correction factor may introduce systematic artifacts or distortions in the spatial distribution, potentially leading to misinterpretation of boron localization. Overall, these findings emphasize that both the acquisition procedure and the determination of correction factors are integral components of the reconstruction process. Accurate boron imaging therefore requires not only precise measurements but also a rigorous and reproducible background characterization, as well as careful validation of the correction factors applied at the pixel level.

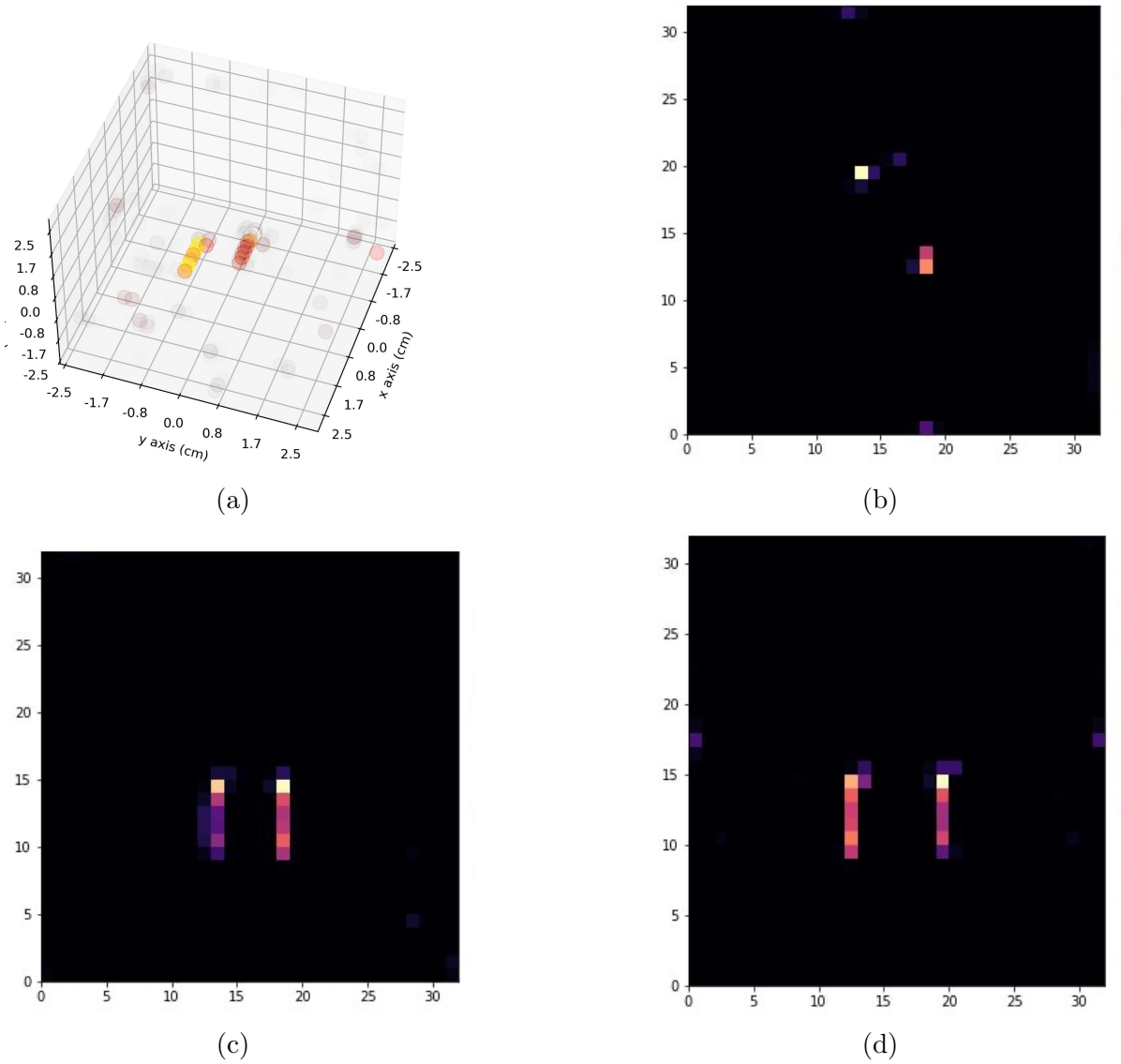


Figure 4.23: Results without correction factors: (a) the reconstruction in 3d; (b) the xz plane projection; (c) the yz plane projection; (d) the xy plane projection.

4.5 Deep learning approach

The analysis and processing of volumetric data, such as 3D medical images or other spatial datasets, often relies on sophisticated deep learning models. These architectures are primarily built upon two fundamental neural network structures: Autoencoders and the U-Net architecture. These models are designed for tasks involving dimensionality reduction, feature extraction, and high-fidelity reconstruction or segmentation.

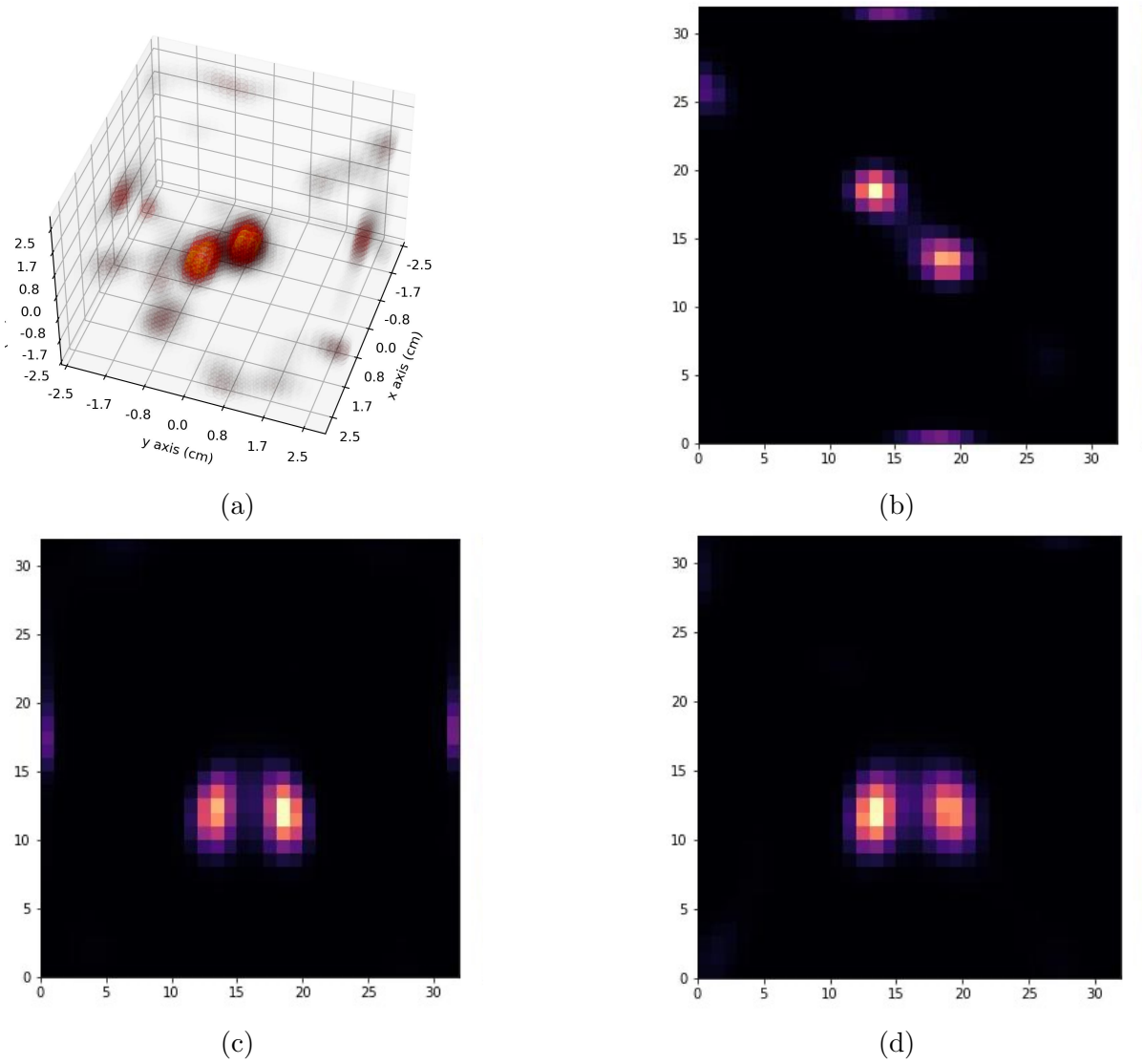


Figure 4.24: Results with correction factors and 20 iterations: (a) the reconstruction in 3d; (b) the xz plane projection; (c) the yz plane projection; (d) the xy plane projection.

4.5.1 Augmentation

In the database are obtained 9 different objects. The way to create a vast database for this type of problem is to augment it by transformations, such rotations, translations, reflections, composition of affine transformations. The noise is add at end of the process. Augmentation pre-processing is applied over all the 9 objects, creating 100 samples.
Augmented images contain an issue: the values in points in which the object doesn't exist are zero, which could compromise the overall outcome of the problem. In order to overcome it, two ways are found: make a reduction of the pixel size, resizing image from

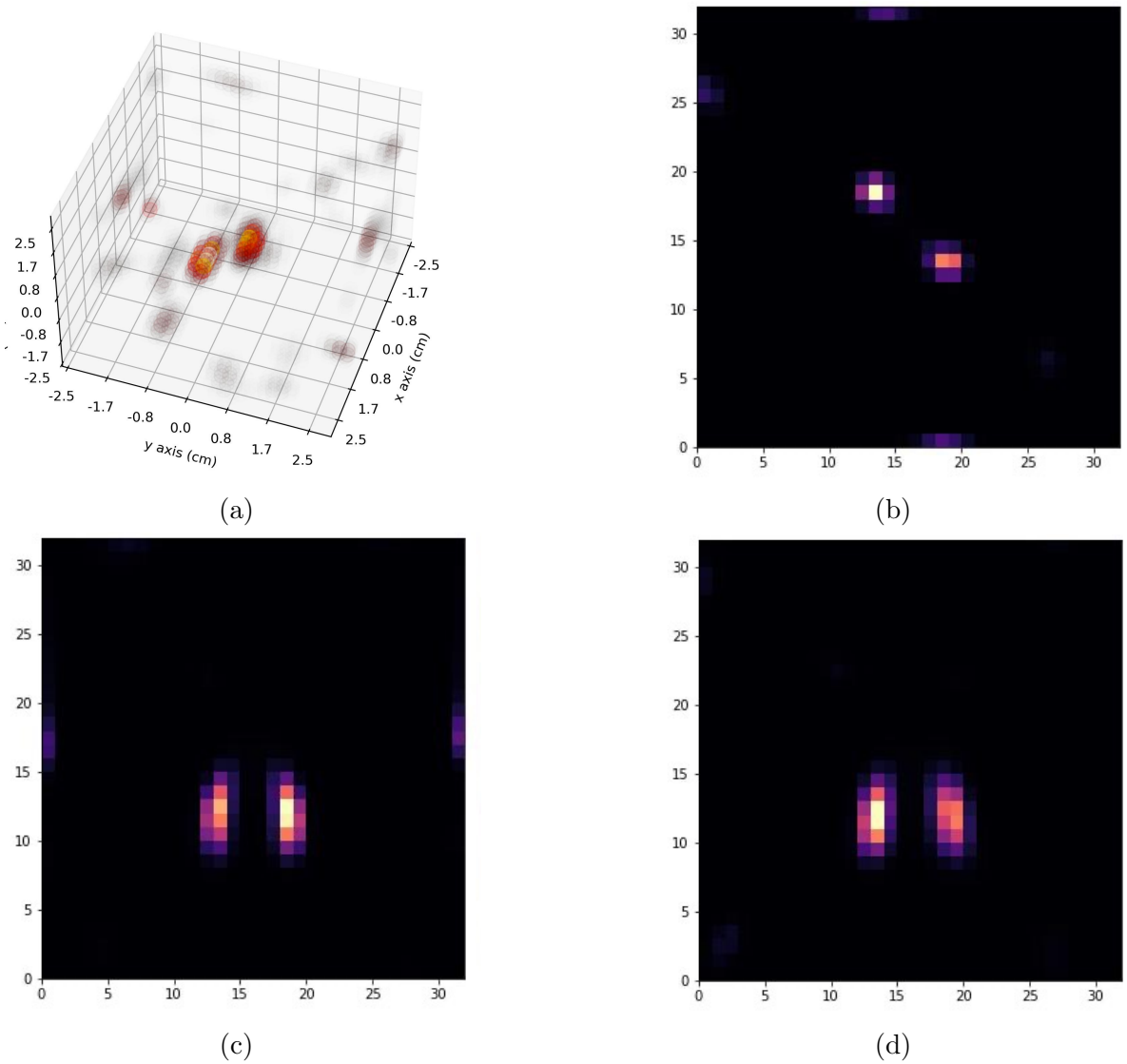
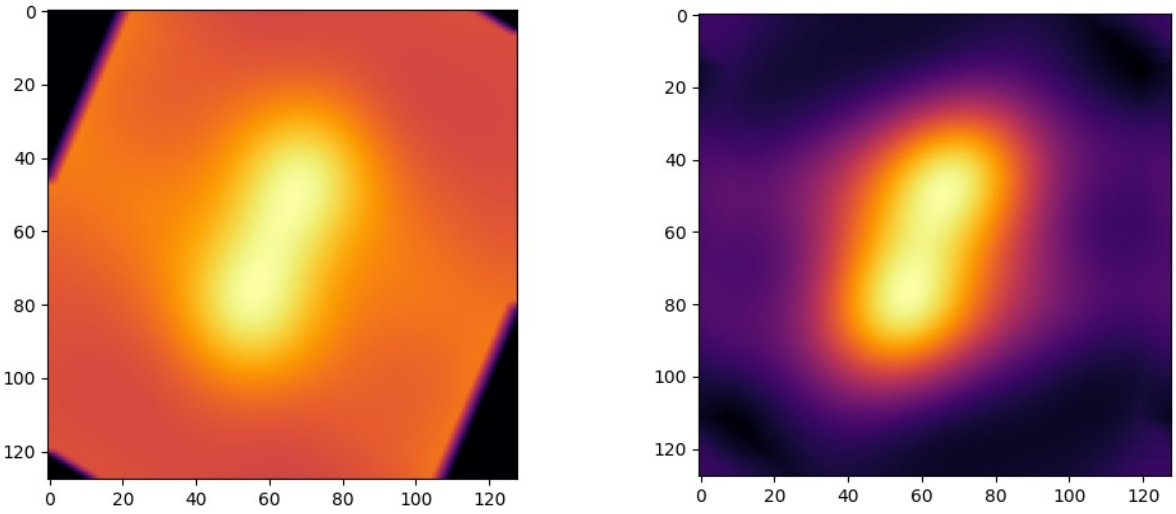


Figure 4.25: Results with correction factors and 20 iterations: (a) the reconstruction in 3d; (b) the xz plane projection; (c) the yz plane projection; (d) the xy plane projection.

128 x 128 x 128 to 64 x 64 x 64, apply the transformation and cut the image to 32 x 32 x 32 with the disadvantage to lose image resolution. The second approach consists of replacing the zero-valued regions introduced by the transformation with background values derived from the object itself, an operation commonly referred to as *blurring*. In this context, blurring can be interpreted as a convolution process in which the original image is locally averaged using a spatial kernel, such as a Gaussian or low-pass filter, that redistributes intensity from neighboring pixels into regions lacking information. Rather than introducing artificial or constant padding, the missing values are estimated from surrounding structures, thereby preserving the global intensity characteristics of the

object and avoiding sharp discontinuities at the boundaries created by the transformation.

Operationally, the blurring process applies a spatially invariant or spatially adaptive kernel whose width determines the extent of information spreading. A narrow kernel limits the influence to immediate neighbors, while a broader kernel propagates intensity over larger regions. Because the kernel operates on the existing pixel grid without resampling, the number of pixels and their spatial spacing remain unchanged, and therefore the nominal spatial resolution of the initial image is preserved. However, while the pixel resolution is maintained, high-frequency image components are attenuated, leading to smoother transitions and reduced local contrast. In tomographic reconstruction and image processing, this form of blurring is often employed as a regularization mechanism, trading fine detail for increased stability and reduced artifacts in regions affected by missing or undefined values. The Fig. 4.26 can show how the process is performed over a test backprojection image.



(a) (b)

Figure 4.26: The result of the blurring process: (a) a unblurred backprojection; (b) the blurring process applied resulting in a visible object, the image is rescaled.

4.5.2 Input Data Specifications

The selection of input volume size is a critical preprocessing step, dependent on both the nature of the data and the computational constraints.

- **Input: $32 \times 32 \times 32 \times 1$ (Preprocessing with Cut):** This smaller input dimen-

sion is utilized when the volumetric data is preprocessed via cutting (or patching). This approach is often necessary for processing extremely large 3D volumes (e.g., high-resolution medical scans) that cannot fit into GPU memory entirely. The small cube of $32 \times 32 \times 32$ pixels focuses the network on local features.

- **Input: $128 \times 128 \times 128 \times 1$ (Preprocessing with Blurring):** This larger input dimension is used when the data is preprocessed using blurring or other global context-preserving techniques, such as downsampling the entire volume. A $128 \times 128 \times 128$ volume allows the network to capture a significantly broader field of view, making it suitable for tasks that require greater global contextual understanding, such as identifying larger structures or general image characteristics.

4.5.3 Training

The training of models is performed using a custom loss function,

$$L = \frac{1}{2}(1 - Dice) + \frac{1}{2}MSE \tag{4.17}$$

where $Dice$ is the coefficient that is defined for target f^* and output $\hat{f}$ as

$$Dice = \frac{2 \times H(\mu_{\hat{f}}, \hat{f}) * H(\mu_{f^*}, f^*)}{\sum_{ijk}^{MNO} H(\mu_{\hat{f}}, \hat{f}) + \sum_{ijk}^{MNO} H(\mu_{f^*}, f^*)} \tag{4.18}$$

and MSE means Mean Squared Error, defined as:

$$\text{MSE} = \frac{\|f^* - \hat{f}\|_2^2}{NMO} \tag{4.19}$$

In that mathematical formulation, $\hat{f}$ denotes the predicted image, while f^* represents the corresponding ground-truth label. The image dimensions are given by M, N, and O, which correspond to the number of pixels along the x, y, and z directions, respectively. The quantities $\mu_{\hat{f}}$ and μ_{f^*} indicate the mean intensity values of the predicted image and the label image. The function $H(\mu_x, x)$ denotes a sigmoid function shifted so that its center is located at the mean value of the variable x. The statistical dispersion of the images is characterized by the variances $\sigma_{\hat{f}}^2$ and $\sigma_{f^*}^2$, which correspond to the predicted image and the label image, respectively, while $\sigma_{\hat{f}f^*}$ represents the covariance between $\hat{f}$ and f^*.

The results are evaluated in terms of three different metrics:

IoU is also called the Jaccard index, and is defined as:

$$\text{IoU} = \frac{H(\mu_{\hat{f}}, \hat{f}) * H(\mu_{f^*}, f^*)}{\sum_{ijk}^{MNO} H(\mu_{\hat{f}}, \hat{f}) + \sum_{ijk}^{MNO} H(\mu_{f^*}, f^*) - H(\mu_{\hat{f}}, \hat{f}) * H(\mu_{f^*}, f^*)} \tag{4.20}$$

The IoU range is [0,1].

SSIM is the Structure Similarity Index Metric, defined as:

$$\text{SSIM} = \frac{(2\mu_{\hat{f}}\mu_{f^*} + c_1)(2\sigma_{\hat{f}f^*} + c_2)}{(\mu_{\hat{f}}^2 + \mu_{f^*}^2 + c_1)(\sigma_{\hat{f}}^2 + \sigma_{f^*}^2 + c_2)} \tag{4.21}$$

where c_1 and c_2 are parameters. It quantifies the similarity based on luminance, contrast and structural information. The structural similarity index takes values in the range $[-1, 1]$, where values close to 1 indicate similarity, a value of 0 indicates no similarity, and values close to -1 indicate anti-correlation.

PSNR is the Peak over Signal Noise Ratio, a value that represents the quantity of signal over the amount of noise in the image, is defined as:

$$\text{PSNR} = 10\log_{10}\left(\frac{\|f^*\|_\infty^2}{MSE}\right) = 20\log_{10}\left(\frac{\|f^*\|_\infty}{\|\hat{f} - f^*\|_2} \cdot \sqrt{NMO}\right) \tag{4.22}$$

where MSE denotes the mean square error and

$$\|f^*\|_\infty = |\max_{i,j,k} f^*(i,j,k)| \tag{4.23}$$

The typical PSNR values are in the range [10,40].

4.5.4 Crop preprocessing

The training set is randomly set as 75% of the database, while the validation is the other 25%. Loss functions' best epochs are in Tab. 4.6. The results are shown in Tab. 4.5. In Fig. 4.27 the predictions for every model are reported. . The Autoencoder is not enough to describe the properties of the dose. The variational AE performs better than the simple AE but suffers from the same problems as the AE. The U-Net overcomes the results of AE. The upgraded version of the U-Net tight frame performs the best, showing that the wavelets decomposition captures spatial-frequency features more explicitly than the simple U-Net. The metrics in Tab. 4.5 show that SSIM is a good evaluator for that regression problem. The inference time for the model U-Net tight frame is about 0.1-0.5 ms using a GPU NVIDIA A100. While, the iteration time for the OSEM algorithm is estimated to be 1 ms for a single subset iteration, so for a subset equal to 3 and with 10 iterations, the iteration time is 30 ms.

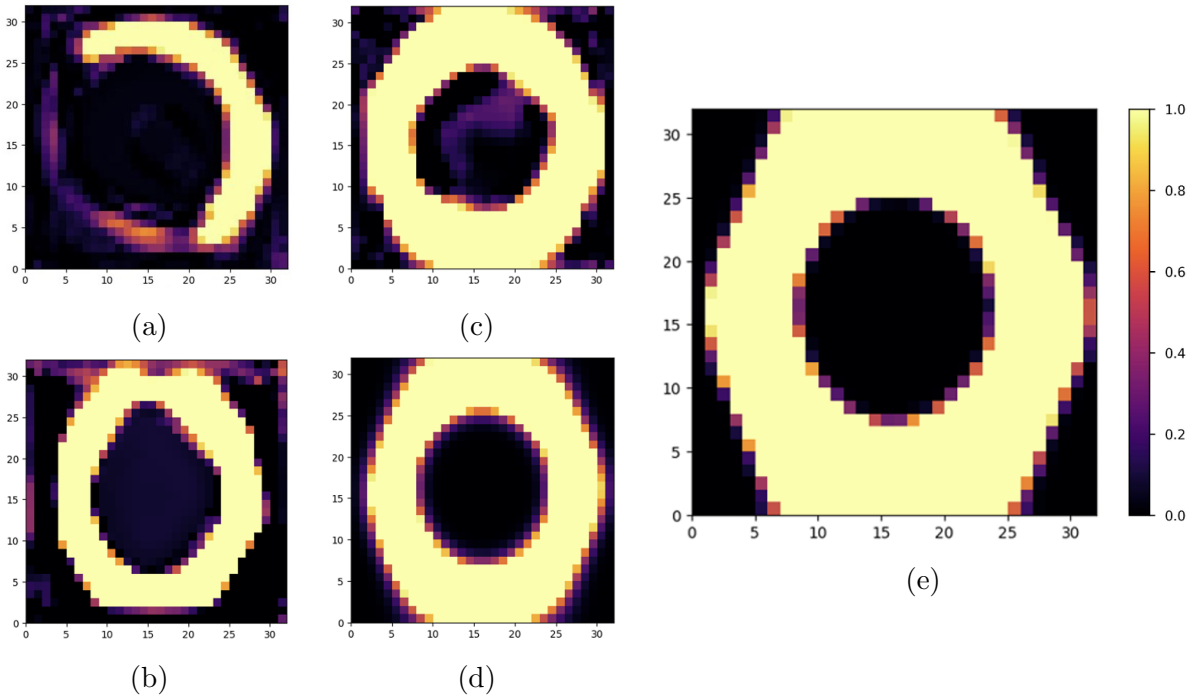


Figure 4.27: Prediction and reconstruction example: (a) AE, (b) VAE, (c) U-Net, (d) U-Net tight frame, (e) reconstructed image.

4.5.5 Blurring preprocessing

Training and validation are the same as the previous preprocessing. Loss functions' best epochs are in Tab. 4.6. The results are shown in Tab. 4.5. In Fig. 4.28 are reported the predictions for every model.

From Tab. 4.5 it is possible to find evidence that after the blurring operation, the SSIM and PSNR for AE are lower than after the cropping operation. This is due to the effect of the larger size of the image. That is not evident for U-Net and U-Net tight frame, because it is overcome by the model performance. Indeed, blurring gives the best results in terms of SSIM and PSNR metrics, with respect to cropping, for the U-Net and U-Net tight frame models. The blurring approach is a non-destructive method. It preserves the information on the geometry and intensity scale. Using the same GPU NVIDIA A100, the inference time is about 5-15 ms. While, the iteration time for the OSEM algorithm is estimated to be 64 ms for a single subset iteration, so for a subset equal to 3 and with 10 iterations, the iteration time is 2-4 seconds. Summarizing, the reconstruction of SPECT tomography is obtained with a deep learning method, which produces excellent results in terms of SSIM and PSNR metrics.

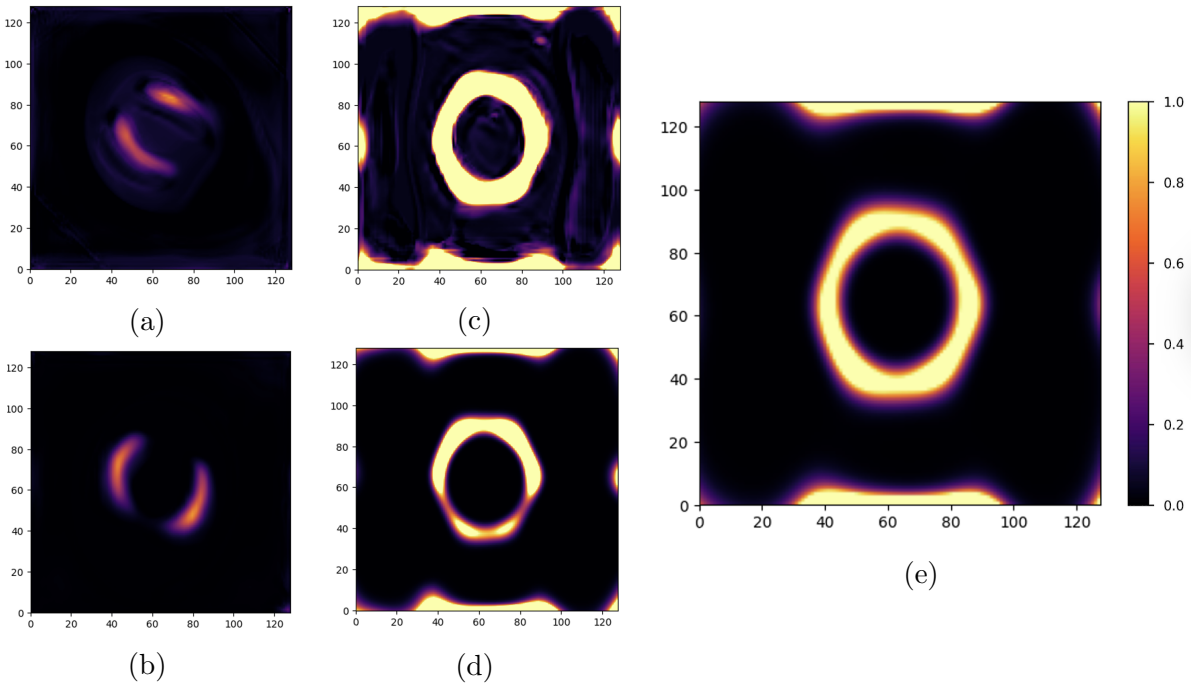


Figure 4.28: Prediction and reconstruction example: (a) AE, (b) VAE, (c) U-Net, (d) U-Net tight frame, (e) reconstructed image.

4.5.6 Discussion

Despite the relatively limited size of the dataset used for training, especially when compared with the large-scale databases typically employed in deep learning applications, the reconstruction results are overall encouraging. This outcome suggests that the chosen model architecture and preprocessing strategy are well aligned with the underlying structure of the reconstruction problem. Nevertheless, the restricted dataset remains a potential limitation, as it may constrain the generalization capability of the network. To mitigate this issue, additional synthetic objects can be generated and incorporated into the training set. Such data augmentation strategies increase variability in object morphology and intensity distributions, thereby improving robustness without requiring new experimental acquisitions. Among the evaluated architectures, the U-Net tight-frame model emerges as the most effective reconstruction framework. Its superior performance can be attributed to the combination of multi-scale feature extraction and the tight-frame formulation, which enforces a stable and redundant signal representation. The encoder–decoder structure of U-Net enables the network to capture both global contextual information and fine spatial details, while skip connections preserve high-frequency features that are critical for accurate reconstruction. The tight-frame constraint further

improves stability by promoting consistent representations across scales, reducing reconstruction artifacts, and enhancing convergence during training. Preprocessing through blurring is shown to play a crucial role in improving reconstruction quality, particularly in the presence of transformed objects containing large regions of zero-valued pixels. These zero values do not correspond to meaningful physical information but instead arise from the transformation process, and if left untreated, they can introduce misleading cues that bias the learning process. The blurring operation addresses this issue by replacing such undefined regions with locally averaged intensity values derived from neighboring pixels. This process preserves the original image grid and nominal spatial resolution while attenuating abrupt discontinuities and suppressing artificial high-frequency components. From a modeling perspective, blurring acts as an implicit regularization step, smoothing the input space and guiding the network toward learning physically meaningful correlations rather than artifacts induced by missing data. As a result, the combined use of a U-Net tight-frame architecture and blurring-based preprocessing leads to more stable training dynamics and improved reconstruction fidelity.

Table 4.5: Results in terms of metric for the two preprocessing methods.

Cropping Model	IoU	SSIM	PSNR
AE	8%	0.45	15
VAE	14%	0.3	19
U-Net	81%	0.86	31
U-Net-TF	85%	0.84	34
Blurring Model	IoU	SSIM	PSNR
AE	15%	0.37	12
VAE	17%	0.38	14
U-Net	88%	0.87	33
U-Net-TF	91%	0.9	38

Table 4.6: Number of training and validation best epochs.

Cropping Model	Best Epoch	Tr. Loss	Vl. Loss
AE	12	0.71	0.88
VAE	17	0.62	0.65
U-Net	18	0.54	0.61
U-Net-TF	21	0.43	0.51
Blurring Model	Best Epoch	Tr. Loss	Vl. Loss
AE	51	0.7	0.95
VAE	28	0.26	0.85
U-Net	47	0.21	0.52
U-Net-TF	53	0.13	0.44

Conclusion and Outlook

This thesis demonstrates that DL can effectively address some of the most persistent challenges in segmentation and particle therapy imaging, offering tangible progress both in diagnostic segmentation and in the reconstruction of the dose for treatment planning. Within the domain of anatomical segmentation, the Two Step strategy developed throughout this work has proven to be a particularly robust response to the difficulties posed by the identification of small and anatomically variable structures such as the pancreas, and the tumour of pancreas. By separating the task of coarse localization by YOLO from the subsequent phase of fine-grained segmentation, the U-Net architecture is able to achieve a high grade of DSC value in defining the tumor boundaries, obtaining a DSC of 58%, that overcomes the state-of-art. This division of tasks allowed the system to cope more effectively with patient's heterogeneity and produced more stable results than conventional One Step pipelines. The One Step study demonstrates that the models with Attention mechanisms are excellent in generalizing the features of CT medical images, because the channel attention mechanism has an architecture shaped to retrieve the features related to the channels. Indeed is demonstrated that the more channels are provided, the more DSC values is high.
Such an approach is essential not only for reducing the workload of clinicians, who traditionally perform these tasks manually, but also for enhancing the reproducibility and reliability of the contours used in downstream therapeutic decisions.

In parallel, the investigation of SPECT reconstruction for BNCT dosimetry has highlighted how DL can substantially improve the quality of low-count boron neutron capture therapy images. Models trained exclusively on Monte Carlo–generated projections demonstrated an ability to suppress noise and recover spatial resolution, such U-Net tight frame model, learning the essential features of the imaging system and providing a reconstruction procedure that is markedly faster than classical iterative techniques such as OSEM. The validation of the FLUKA and Geant4 simulation environments ensured

that these synthetic datasets can function as trustworthy digital counterparts of the real apparatus, supporting future research on detector development and on the optimization of BNCT acquisition protocols. The DL pipeline used can be easily reproduced in a real facility with a clinically acceptable neutron flux. However, it must be emphasized that full in-vivo dose reconstruction in BNCT using a tomography setup based on gamma cameras remains a long-term goal. What has been successfully achieved in this framework is the essential groundwork: demonstrating the feasibility and computational efficiency of DL-based reconstruction on experimental and simulated vials.

The use of classical reconstruction is always suggested and mandatory in this set-up, but the model can be simply upgraded taking as input the projections directly coming from detector acquisition, without the use of the backprojection step. It will be investigated in further studies. The results obtained with U-net tight frame, for SSIM and PSNR metrics, demonstrated that with a little modification of the model, it could certainly be trained to overcome this result. BNCT images, both experimental and simulated, will be used as a starting point to build a database big enough to be ingested by modern DL models.

Taken together, these contributions show that integrating machine vision into the medical physics workflow is not only feasible but highly advantageous. Deep Learning can serve as a unifying framework that connects the precision of computational methods with the needs of clinical practice, enabling more accurate tumor localization, more accurate BNCT imaging, and ultimately a pathway toward treatments that are better tailored to the individual patient. By bridging the gap between advanced algorithmic tools and real-world therapeutic applications, this work outlines a future in which diagnostic and therapeutic procedures operate with greater automation, consistency, and precision, improving patient care in fields where early detection and accurate dosimetry remain crucial challenges.

Despite these promising results, transitioning these Deep Learning models from controlled experimental or simulated environments to real clinical systems entails several significant challenges. First, there is the critical issue of domain shift: models trained predominantly on simulated (Monte Carlo) or idealized phantom data often experience performance degradation when applied to real patient anatomy. This is due to unmodeled complexities such as heterogeneous tissue attenuation, patient motion, and highly variable background radiation. In real BNCT applications, the background gamma spectrum (e.g.,

from hydrogen neutron capture) is highly dependent on patient-specific factors, making the synthesis of perfectly accurate training data extremely difficult. Second, data scarcity and annotation remain major bottlenecks. Training robust DL architectures requires massive datasets with precise ground-truth labels, which are notoriously difficult and time-consuming to obtain in rare and emerging modalities like BNCT-SPECT. Furthermore, the black-box nature of neural networks poses a critical challenge for clinical adoption. In radiotherapy dosimetry, a network hallucinating a signal or suppressing a true anatomical feature could lead to severe miscalculations of the delivered dose. Therefore, future efforts must focus not only on scaling the datasets but also on developing explainable AI methods, uncertainty quantification, and rigorous physics-informed regularization to ensure that the network's outputs are not only accurate but also clinically trustworthy and robust against real-world artifacts.

To ensure the reproducibility of the simulations and experiments presented in this thesis, the primary data and methodologies are documented. The datasets utilized in the pancreatic segmentation study (MSD and CPTAC-PDA) are publicly available through their respective online repositories. The proprietary Geant4 and FLUKA simulation data, as well as the raw experimental measurements acquired from the LENA facility, are available from the corresponding authors upon reasonable request. The custom code and scripts developed for model training, data preprocessing, and tomographic reconstruction implemented primarily using TensorFlow, PyTorch, and PyTomography will be made available in a public repository upon the formal publication of this work (or are available upon reasonable request) to allow full replication of the analytical pipelines.

Appendix A

Study of the radiation fields at CERN Gamma Irradiation Facility

A.1 Introduction

The Gamma Irradiation Facility (GIF++) is an experimental area at CERN where a Cs-137 source and a high-energy muon beam can be combined for long-term gaseous detector studies. The facility was built in 2014 during the Long Shutdown 1 (LS1) of the LHC and became operational in the spring of 2015 [119], succeeding the original GIF facility that was in use from 1997 to 2014.

Today, GIF++ plays a crucial role in testing muon detectors under high-background radiation conditions, particularly in preparation for the High-Luminosity phase of the LHC (HL-LHC). Furthermore, GIF++ will provide an unique infrastructure for various R&D initiatives under the umbrella of the recently established DRD1 Collaboration.

Currently, the facility hosts detectors such as Resistive Plate Chambers (RPCs), Cathode Strip Chambers (CSCs), and Micromegas (MMs), opereted by several experiments (CMS, ATLAS, ALICE, LHCb, TOTEM, and SHIP). Additionally, GIF++ serves as the experimental headquarters of the RPC EcoGas@GIF++ Collaboration, focuses on the study of environmentally friendly gas mixtures as replacements for traditional RPC operation.

In 2018, the facility underwent an expansion to enhance its capabilities in the upstream region, along with other minor modifications. To accommodate these changes, an updated

simulation is required to accurately reflect the new geometry.

This paper aims to provide a comprehensive characterization of the gamma fields at GIF++ under different irradiation conditions in the experimental areas. The study continues some previous work conducted in 2017 [120]. Measured dose rates are compared with simulations to improve the understanding of the radiation field and validate the simulation framework.

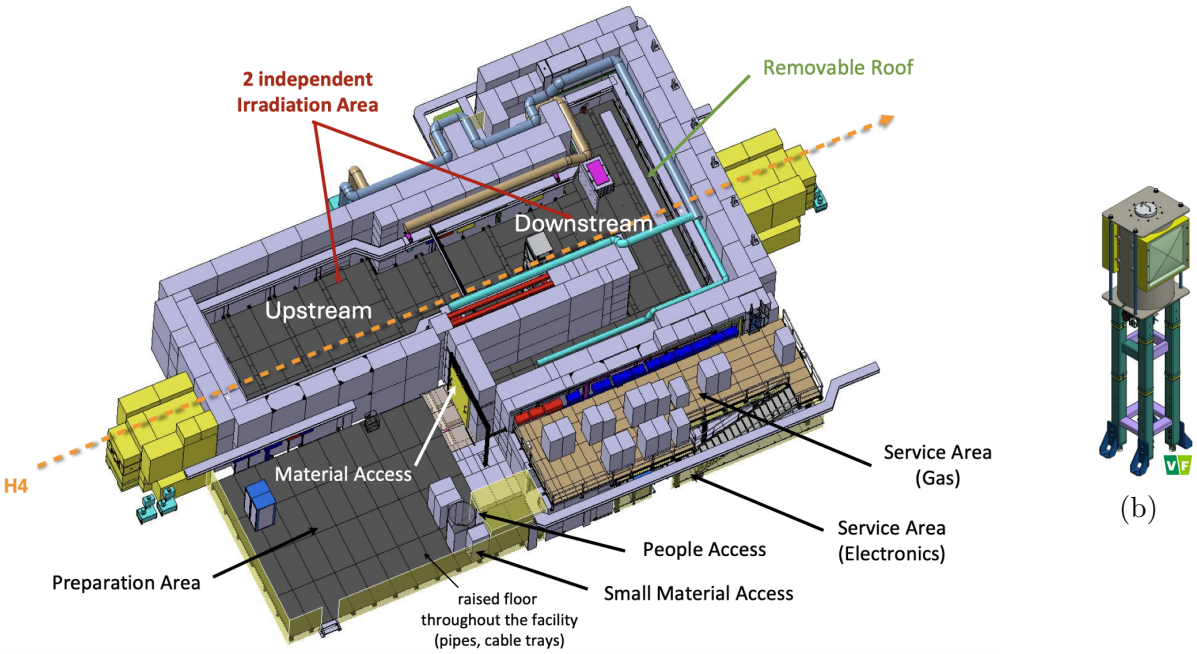
A.2 The Gamma Irradiation Facility

GIF++ is located at the H4 beamline in CERN's Prévessin North Area. The facility provides a no monochromatic muon beam with a momentum typically in the range 10 - 150 GeV/c [121] and a high-intensity Cs-137 gamma source, initially in March 2015 with an activity of 13.9 TBq. At the time of the measurement reported in this work (spring 2025) the activity was 11 TBq. A 3D model of the facility is shown in Fig. A.1a, highlighting three main areas: the irradiation bunker (top, with the beam entering from the left), the preparation area (bottom left), and the gas and rack area (bottom right), which supplies gas and electronic services to the bunker [122].

The irradiator (Fig. A.1b) was developed in collaboration with VF Nuclear[1]. The cesium source, enclosed between two tungsten blocks, moves between a ground position at the bottom of a support tube and an irradiation position at the top. Two $\pm$ 37° panoramic collimators define the irradiation zone, covering a large part of the bunker, which is divided into upstream (UP) and downstream (DOWN) areas, named according to the beam direction. Mounted on an adjustable support, the center of the irradiation window aligns with the beam height, delivering up to $\approx$1 Gy/h at a 1 m distance. Both collimators include replaceable steel angular correction filters to ensure uniform photon distribution for large-area detectors.

The irradiator also features two independent attenuation systems housed within a common enclosure, each with a 3×3 array of absorption filters to fine-tune the photon flux in each irradiation field. The total attenuation factor combines three layers (A, B, C) and can be adjusted using the attenuation matrix in Tab. A.2, either via a dedicated control panel or the GIF++ control system. A total of 24 nominal attenuation factors (ABS), ranging from 1 to 46420, can be selected from 27 possible configurations (Tab. A.2).

[1]VF Nuclear, Czech Republic, http://www.vfnuclear.com



(a) (b)

Figure A.1: (a) 3D model of GIF++ experimental area, (b) view of the irradiator [119]

.

When ABS is 1, the source is fully open, meaning photons pass through the angular correction filter without additional attenuation. At the maximum ABS value of 46420, the source is considered fully closed.
The convex filters are made of lead coated with protective paint. Mounted on aluminum support plates, they are housed within steel frames and connected to counterweights on the side of the irradiator. The list of the filters is in Tab. A.1.

The detailed geometric layout is shown in Fig. A.2. The red points evidenced in the figure show some defined positions both in the upstream area (U1, U2, U3, U3R, U3L, U5, U5R, U6, U7, U11) and in the downstream area (D1, D3, D5) where dedicated measurements were taken or dedicated simulations were performed. In Tab. A.3 the XYZ coordinates of such positions are given.

A.3 Simulation methods

As previously mentioned, the main scope of this study is to simulate dose and gamma current in the bunker area taking into account the latest modifications applied to the facility. The new bunker geometry has an extension of 14.7 m from the source to upstream

Table A.1: Attenuation factors matrix for filter combinations

	A	**B**	**C**
1	1	1	1
2	10	1.468	2.154
3	100	100	4.642

Table A.2: Filter combinations and corresponding approximated ABS values

Filter Combination (ABC)	**ABS**
111	1
112	2.2
122	3.3
113	4.6
123	6.9
211	10
212	22
222	33
213	46
223	69
311	100
333	46420

	U1	U2	U3	U3R	U3L	U5	U5R	U6	U7	U11	D1	D3	D5
x	0	0	0	+1.6	-1.6	0	+2.4	0	0	0	0	0	0
y	0	0	0	0	0	0	0	0	0	0	0	0	0
z	-1	-2	-3	-3	-3	-5	-5	-6	-7	-11	+1	+3	+5

Table A.3: Table of positions from the source of measurements and simulation points in meter. U11 is used only for simulations.

and 6.7 m downstream along the main axis. The simulation was performed in Geant4 exploding version 11.3 [116] and the G4EmLivermore Physics physics list. The geometric layout shown in Fig. A.2 was implemented.

Geant4 allows to upload a Geometry Description Markup Language file (GDML) in which the geometry and materials are modelled. A GDML file was developed for each ABS value in Tab. A.2 The whole facility was partitioned by using meshes of $5 \times 5 \times 5$ cm^3 as scoring cubes to retrieve absorbed dose and photon currents. The absorbed dose rate [Gy/h] was calculated from the simulated dose [Gy] by using a reference time following the formula A.1.

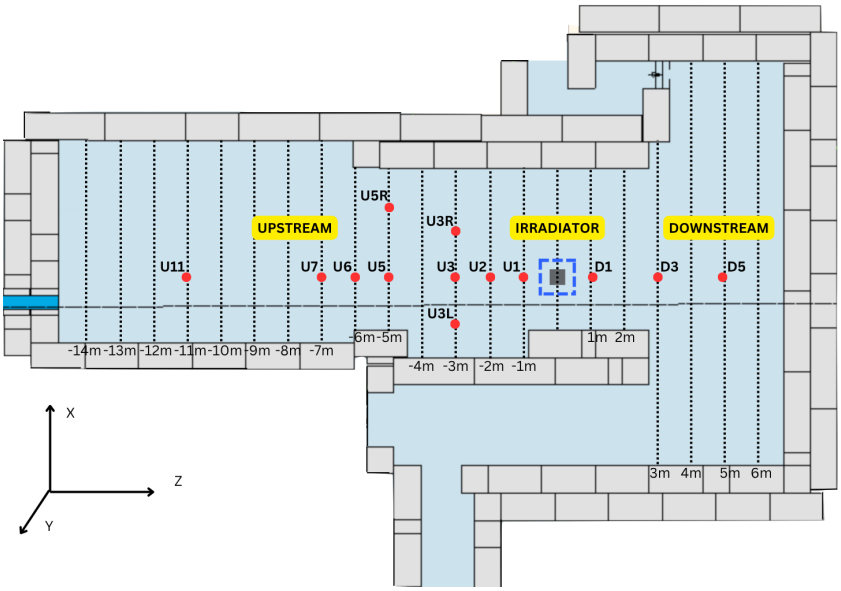


Figure A.2: Layout of the GIF++ experimental area indicating the dose measurement positions. The reference system in shown. The point (0,0,0) is located at the source position.

$$t^s = \frac{N_p}{BF \cdot A \cdot 3600} \tag{A.1}$$

In the Eq. A.1, N_p is the number of photons ejected from the source in a simulation run, BF is the branching fraction of ^{137}Cs decay schema, set to 0.85 [123], while A represents the source activity.

The errors in the simulation were calculated following Eq. A.2 by accounting only the statistical uncertainty following the Monte Carlo approach.

$$u(D_{MC})/\overline{D}_{MC} = \frac{\sqrt{\sum_{i=1}^{N_{event}} q_i^2}}{\sum_{i=1}^{N_{event}} q_i} \tag{A.2}$$

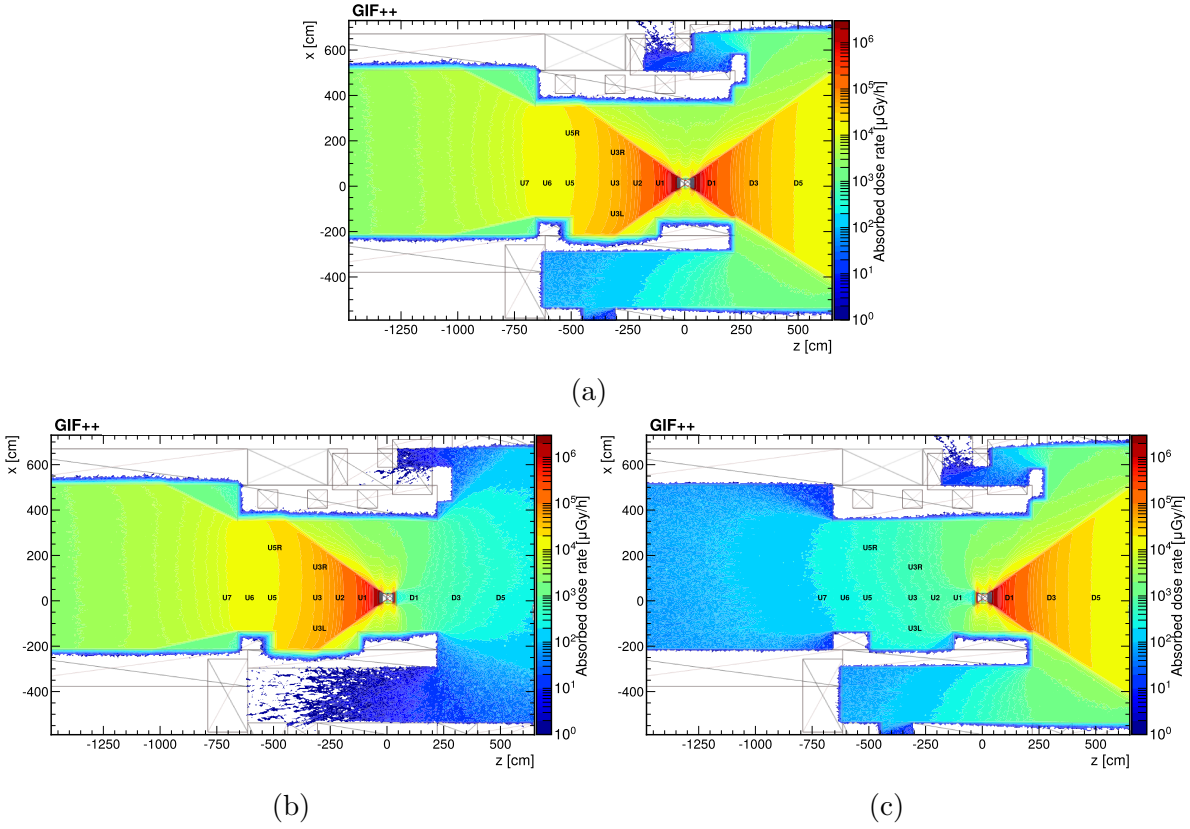
where D_{MC} is the dose calculated in every volumes, and q_i is the dose for each event.

A.4 Simulation results

Fig. A.3 shows the total absorbed dose rate (XZ view) with the source fully open both in the upstream and downstream regions. A slightly concave dose profile is observed along

the x-axis, more pronounced at greater distances from the gamma source. However, the profile appears relatively flat for points within $|x| \leq 1.5$ m.

A possible back-scattering effect was studied by simulating a full irradiation in one region and maximum attenuation in the opposite region. Figures A.3b and A.3c show the XZ dose profiles respectively for the cases where the upstream field remains open while the downstream is closed, and vice versa. No noticeable back scattering effect is evident.



(a)

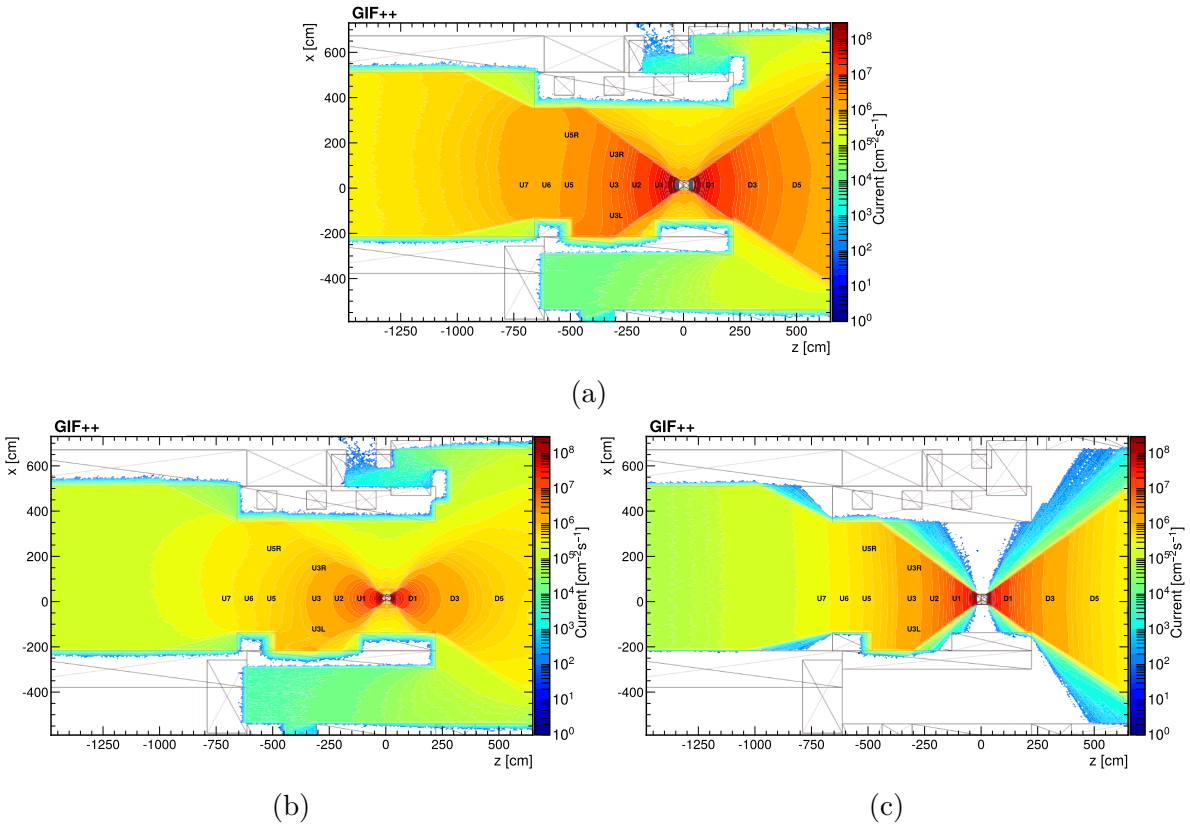
(b) (c)

Figure A.3: XZ-view of total absorbed dose rate maps over the irradiation bunker when (a) setting ABS 1 upstream and ABS 1 downstream, (b) setting ABS 1 upstream while downstream is fully closed, and (c) setting ABS 1 downstream while upstream is fully closed.

A.5 Gamma spectral analysis

The gamma energy spectrum was also simulated, being relevant for the sensitivity study of the detectors under test. Fig. A.4 shows the XZ total gamma flux with maximum irradiation in both upstream and downstream regions for the three different cases: A.4a all gammas energies, A.4b gamma energies below 600 keV, A.4c gamma energies over

600 keV. It can be noticed that the gamma flux at the cesium peaks is uniform along X while the low energy components tends to be more pronounced at small X.



(a)

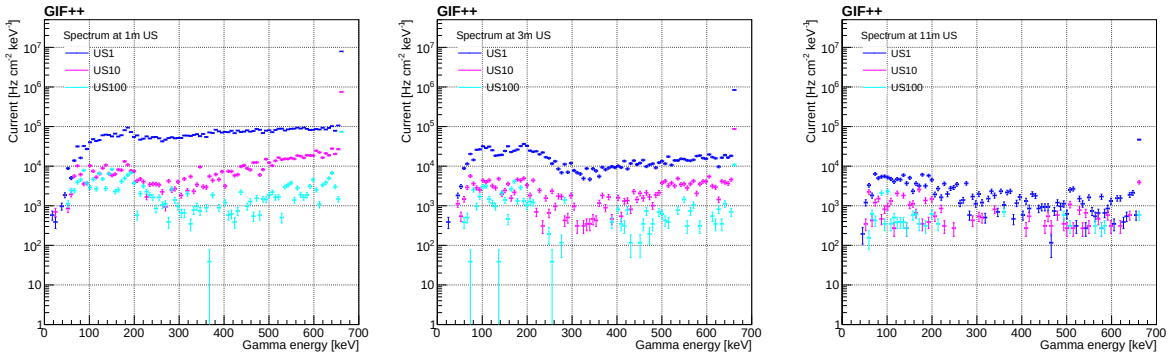
(b)

(c)

Figure A.4: Map of gamma flux for (a) all energies, (b) energy range 0-600 keV, (c) energy higher than 600 keV.

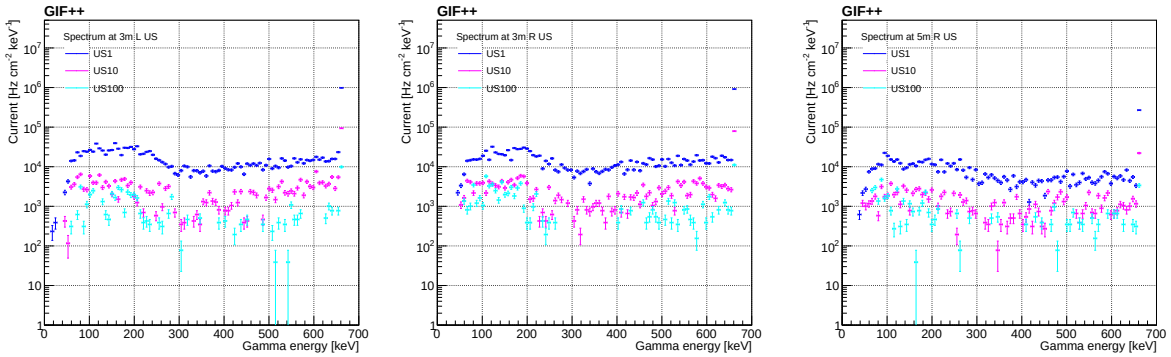
Gamma spectral analysis was also performed at positions U1, U3, U11, D1, D3, and D5, considering different attenuation factors in the upstream region (ABSs 1, 10, 100) and constant attenuation in the downstream region (ABS 2.2). The energy spectra for the upstream positions are shown in Fig. A.5, while detailed gamma fluxes for all positions are presented in Tab. A.4.

In the spectra reported in Fig. A.5, a reduction in the energetic component associated with the 662 keV peak is observed for all ABSs. It is confirmed that at increasing values of z the spectrum is dominated by low energy gammas.



(a) (b) (c)

Figure A.5: Gamma energy distribution at positions (a) U1, (b) U3, (c) U11. Different attenuation factors are considered in the upstream region (ABSs 1, 10, 100), while a constant attenuation is considered in the downstream region (ABS 2.2). The histograms present a resolution of 7 keV/bin.



(a) (b) (c)

Figure A.6: Gamma energy distribution for the positions (a) U3L, (b) U3R, (c) U5R. Different attenuation factors are considered in the upstream region (ABSs 1, 10, 100), while a constant attenuation is considered in the downstream region (ABS 2.2). The histograms present a resolution of 7 keV/bin.

A.6 Dose measurements

The dose measurement campaign was conducted in the GIF++ irradiation bunker during November 2024. A gamma probe 6150AD-15 model H, coupled with a dose rate meter [124], was employed. The probe detector features an integrated Geiger-Müller counter. The useful dose rate range of the probe is 1 mSv/h to 9.99 Sv/h. The probe was calibrated by the manufacturer with a 333 kBq Cs-137 calibration source. Within the useful dose rate range, a maximum deviation of ± 10% in linearity between the measured dose rate

Table A.4: Gamma flux values corresponding percentages for different energy ranges.

	Flux $[10^6 cm^{-2} \cdot s^{-1}]$					%				
ABS	Total	0-200 keV	200-400 keV	400-600 keV	600-663 keV	0-200 keV	200-400 keV	400-600 keV	0-600 keV	600-663 keV
U1										
1	40.92	3.05	4.96	6.95	25.96	7	12	17	37	63
10	4.66	0.45	0.34	1.06	2.81	10	7	23	40	60
100	0.83	0.28	0.10	0.14	0.30	34	12	17	63	37
U3										
1	6.58	1.51	1.01	1.12	2.93	23	15	17	55	45
10	0.89	0.22	0.10	0.22	0.35	24	11	25	60	40
100	0.19	0.10	0.01	0.03	0.04	52	7	17	77	23
U11										
1	0.70	0.28	0.16	0.09	0.17	40	23	13	76	24
10	0.12	0.06	0.02	0.02	0.01	54	14	20	88	12
100	0.04	0.03	0.00	0.01	0.00	76	7	13	96	4
D1										
1	41.28	3.14	4.97	7.02	26.17	8	12	17	37	63
10	4.78	0.47	0.35	1.08	2.88	10	7	23	40	60
100	0.86	0.31	0.10	0.16	0.30	36	11	18	65	35
D3										
1	6.21	1.23	0.99	1.15	2.85	20	16	18	54	46
10	0.88	0.19	0.10	0.22	0.37	22	11	25	58	42
100	0.19	0.10	0.02	0.03	0.04	54	10	17	81	19
D5										
1	2.78	0.91	0.56	0.41	0.90	33	20	15	68	32
10	0.41	0.13	0.06	0.10	0.11	32	15	25	72	28
100	0.09	0.05	0.02	0.01	0.01	59	17	17	93	7

and the calibration is permitted. The energy range spans from 65 keV to 3 MeV, and the nominal angular range extends ±45° relative to the preferred direction perpendicular to the axis of the tube. The position of the probe was consistently maintained at a right angle, always directed towards the source. During the measurements, it was ensured that no other material (detectors under test or mechanical supports) were in the path between the source and the dosimeter.

The gamma probe measurements were performed at all positions (but U11) defined in Fig. A.2 for several ABS values (1, 2.2, 3.3, 4.6, 6.9, 10, 22, 33, 46, 69, and 100). For the upstream measurements the downstream ABS value was kept constant to 2.2, for the downstream measurements the upstream ABS value was kept constant to 2.2.
Three minutes of data were collected for each filter configuration to ensure enough statistics. The gamma probe response is the Ambient Equivalent Dose rate ($H^*(10)$[Sv/h]). Based on the technical sheet specifications, the overall dose measurement uncertainty

was assumed to be ±40%.

The Absorbed Dose rates values (D[Gy/h]), needed for comparison with the simulated values, were obtained by the Eq. A.3 where $f2(E)$ for the Cs-137 gammas was assumed equal to 1.20[2].

$$D = H^*(10)/f2(E) \tag{A.3}$$

A.7 Simulation validation

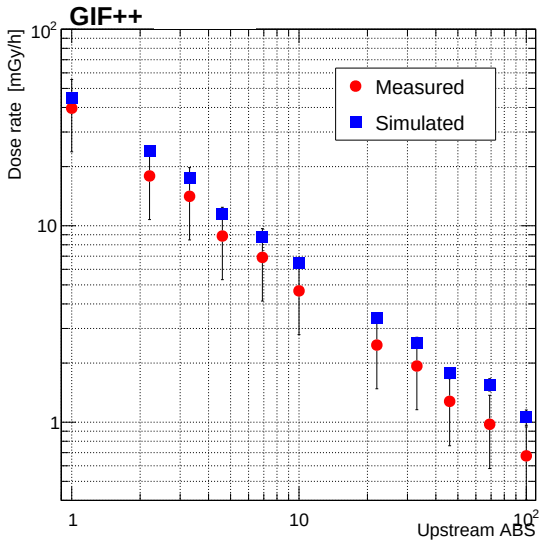


Figure A.7: Measurements and simulations at different ABS values in the upstream region (keeping constant ABS = 2.2 in the downstream region) at position U3.

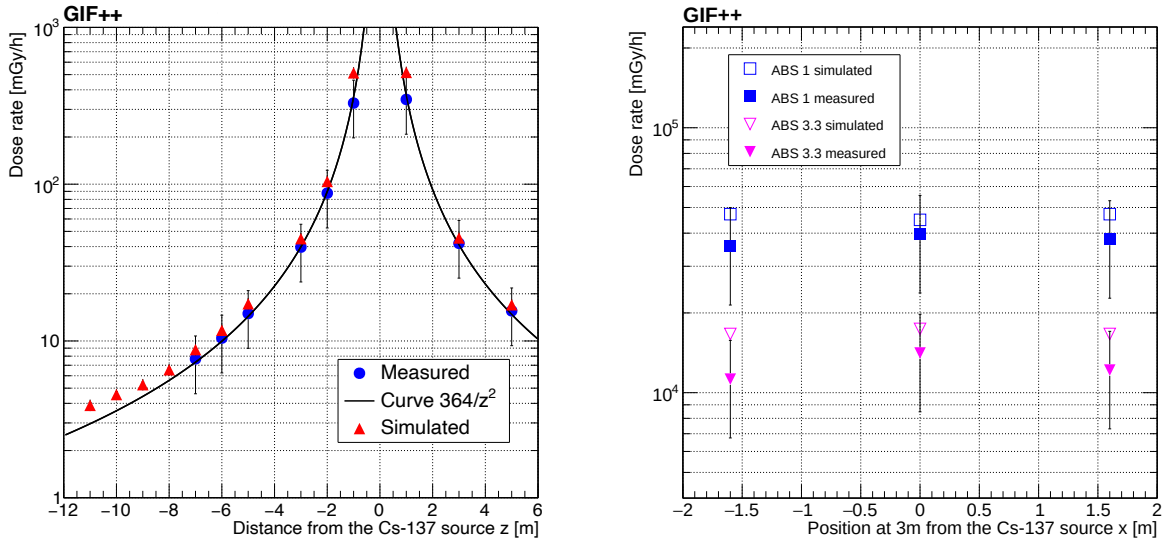
In Fig A.7, measurements with the gamma probe are compared to the simulation results at different ABS values in the upstream region (keeping constant ABS = 2.2 in the downstream region) at position U3. All data at x = 0 is fitted following the proportionality of the dose to the square inverse of the distance from the source: $D = b/z^2$, where D is Absorbed Dose rate, z is the distance from the source on beam axis at x = 0 and b is a free parameter. Figure A.8a reports the measured and simulated data together with the fitted function for the case of ABS = 1 in both irradiation fields. Data and simulation are quite in good agreement for high values of $|z|$. At $|z| = 1$, although still within the error bars, the simulation value is slightly higher, perhaps due to the

[2]Automess - Radiation Quantities and Units, https://www.automess.de/en/service/radiation-quantities-and-unitsphoton-dose-equivalent-hx

Table A.5: Results of the fit for the parameter b, at different ABS values.

ABS	1.0	2.2	3.3	4.6	6.9	10	22	33	46	69	100
b	364	164	128	81	62	42	21	16	10	8	5

detector angular acceptance which introduces a larger bias in the measurements at short distance from the source. The fit accurately describes the data and it can be used for dose estimation throughout the entire bunker. Results of the fit for the parameter b, at different ABS values (assuming equal attenuation for both fields), are shown in Tab. A.5.



(a) (b)

Figure A.8: Absorbed dose rate as a function of the (a) distance from the source in x = 0, (b) position in x-axis at 3 m from the source in the upstream area.

The dose study along the x-axis was carried out at z = 3 m from the source on the upstream field. The positions used for the analysis were U3L, U3, and U3R. The dose values for ABSs 1 and 3.3 are shown in Fig. A.8b. Agreement between simulation and measurements is good at x = 0 (U3), while off-axis (U3R and U3L), the experimental values are slightly lower than the simulated ones. This behavior might be due to the fact that the gamma probe is calibrated for Cs-137 energy, while in the off-axis region also a low-energy component flux is present as discussed in Sec. A.5. The ratio between simulation and measured dose rate values is reported in Tab. A.6.

Table A.6: Values of ratio between simulation and measure for all position and ABS.

ABS	U1	U2	U3	U3R	U3L	U5	U5R	U6	U7	D1	D3	D5
1	1.56	1.18	1.13	1.24	1.32	1.15	3.07	1.12	1.15	1.48	1.08	1.09
2.2	2.22	1.42	1.34	1.46	1.56	1.22	3.66	1.35	1.27	2.10	1.20	1.23
3.3	2.07	1.40	1.23	1.39	1.48	1.25	3.39	1.33	1.20	1.95	1.15	1.14
4.6	2.37	1.48	1.30	1.62	1.75	1.36	3.60	1.48	1.33	2.19	1.32	1.26
6.9	2.24	1.44	1.26	1.57	1.72	1.42	3.14	1.21	1.29	2.05	1.20	1.25
10	2.26	1.51	1.39	1.64	1.83	1.24	3.68	1.45	1.34	2.15	1.42	1.42
22	2.36	1.43	1.38	1.84	1.96	1.68	3.99	1.67	1.17	2.14	1.28	1.44
33	2.13	1.44	1.32	1.61	1.79	1.27	3.90	1.38	1.20	1.96	1.28	1.43
46	2.42	1.67	1.40	1.67	1.81	1.64	3.09	1.20	2.23	2.10	1.47	1.49
69	2.27	1.27	1.59	1.62	1.82	1.48	3.93	1.35	1.52	1.98	1.34	1.32
100	2.39	1.62	1.57	2.12	2.30	1.67	4.47	1.49	1.83	2.11	1.65	1.90

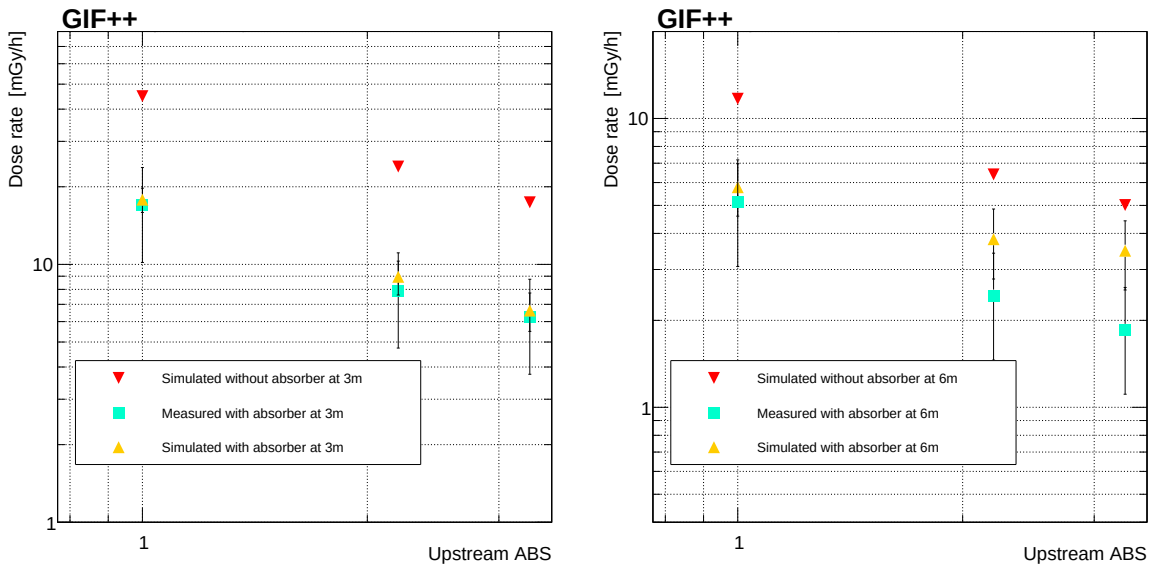
A.8 Dose absorption in operation conditions

To study realistic dose rates during GIF++ operation, two new measurements were taken at positions U3 and U6 in the presence of an additional detector setup that shadows the source. By comparing the measurements with (D_m) and without(D_0) such absorber material, an equivalent Aluminum thickness Δx can be inferred using the equation: [125, 126]:

$$D_m = f D_0 e^{-\mu \Delta x} \tag{A.4}$$

where μ is the attenuation coefficient and f the buildup factor at the Cs-137 energy. The calculation was performed using the online available resource RadProCalculator [127, 128]. The equivalent Aluminum thickness was found to be $\Delta x = 8.7$ cm at position U3 and $\Delta x = 8.1$ cm at position U6.

The smaller equivalent thickness value obtained at U6 compared to the one at U3 can be attributed to the increased low-energy component at greater distances from the source. Simulated dose rate were therefore recomputed at U3 and U6 considering also the additional shadowing material. Figure A.9 shows the dose rates with and without material for ABS values 1, 2.2, and 3.3. The simulation is in agreement with the measured dose rate inside errors.



(a) (b)

Figure A.9: Comparison between the simulated dose with/without equivalent Al and the experimental data for ABS values 1, 2.2, and 3.3 at (a) 3 m and (b) 6 m from the source in upstream field.

A.9 Conclusions

A detailed characterization of the radiation fields at the CERN Gamma Irradiation Facility has been performed through a combination of dedicated measurements and Geant4-based simulations. Dose rate distributions were obtained across the irradiation bunker under different attenuation configurations and validated against experimental data collected with calibrated gamma probe 6150AD-15. The agreement between measured and simulated values is within the assigned uncertainties, confirming the reliability of the updated simulation framework to reproduce the radiation environment. The spatial dose profiles demonstrate the expected inverse-square dependence with distance, with deviations at larger distances explained by the buildup of scattered radiation. Lateral measurements also highlight the influence of low-energy components, which become more relevant with increased attenuation. The gamma spectral analysis confirms the predominance of the 662 keV Cs-137 line close to the source, while low-energy components progressively dominate further away and under strong attenuation. The introduction of equivalent material thicknesses in the simulations allowed a realistic description of beam test conditions, showing good agreement with measured dose absorption. These results provide an essential benchmark for users operating detectors inside the facility and serve

as a validated input for future studies of radiation effects on detector technologies in high-background conditions.

Bibliography

[1] Bengt Falk, David Strand, and Oscar Larsson.
Early detection of pancreatic cancer: Current advances and future opportunities.
Biomedicines, 13(7):1733, 2025.

[2] Margaret Rosen, Daniel Li, and Mika Ishikawa.
Imaging for pancreas adenocarcinoma: Staging, surgical considerations, and surveillance.
World Journal of Gastrointestinal Surgery, 15(11):2315–2332, 2023.

[3] Scott Adams, Shankar Saini, and Liang Zhu.
Imaging for pancreatic ductal adenocarcinoma.
Seminars in Roentgenology, 53(1):15–28, 2018.

[4] Joel S. Brown, Sarah R. Amend, Robert H. Austin, Robert A. Gatenby, Emma U. Hammarlund, and Kenneth J. Pienta.
Updating the definition of cancer.
Molecular Cancer Research, 21(11):1142–1147, 11 2023.

[5] Anubhav Garg, Shrey Raman, and Jason Lee.
Early detection of pancreatic cancer on computed tomography: Advancements with deep learning.
Radiology Advances, 2024.

[6] Y. Huang, R. Patel, and G. Chen.
Pancreatic adenocarcinoma: Imaging modalities and the role of artificial intelligence in analyzing ct and mri images.
Cancers, 16(3):537, 2024.

[7] Rongmiao Zhou, Zhun You, and Liang Liu.

The basis and advances in clinical application of boron neutron capture therapy. *Radiation Oncology*, 16, 2021.

[8] Timothy D. Malouff, Danushka S. Seneviratne, Daniel K. Ebner, William C. Stross, Mark R. Waddle, Daniel M. Trifiletti, and Sunil Krishnan.
Boron neutron capture therapy: A review of clinical applications.
Frontiers in Oncology, 11:601820, 2021.

[9] Simon Cherry, J.A. Sorenson, and Michael Phelps.
Physics in Nuclear Medicine, volume 31, page 523.
01 2003.

[10] Luca Ceriani, Teresa Ruberto, Angelika Bischof Delaloye, John O. Prior, and Luca Giovanella.
Three-dimensional ordered-subset expectation maximization iterative protocol for evaluation of left ventricular volumes and function by quantitative gated spect: A dynamic phantom study.
Journal of Nuclear Medicine Technology, 38(1):18–23, 2010.

[11] Lai K. Leong, Randall L. Kruger, and Michael K. O'Connor.
A comparison of the uniformity requirements for spect image reconstruction using fbp and osem techniques.
Journal of Nuclear Medicine Technology, 29(2):79–83, 2001.

[12] R. Schofield, L. King, U. Tayal, I. Castellano, J. Stirrup, F. Pontana, J. Earls, and E. Nicol.
Image reconstruction: Part 1 – understanding filtered back projection, noise and image acquisition.
Journal of Cardiovascular Computed Tomography, 14(3):219–225, 2020.

[13] Sarit Ben-Haim et al.
Pet myocardial perfusion imaging.
Journal of Nuclear Cardiology, 20(5):972–978, 2013.

[14] Maria R. Ponisio, John M. Zempel, Brian K. Day, Lawrence N. Eisenman, Michelle M. Miller-Thomas, Matthew D. Smyth, and R. Edward Hogan.
The role of spect and pet in epilepsy.
American Journal of Roentgenology, 216(3):759–768, 2021.

[15] Johann Radon.
Über die bestimmung von funktionen durch ihre integralwerte längs gewisser mannigfaltigkeiten.
Berichte über die Verhandlungen der Sächsischen Akademie der Wissenschaften, 69:262–277, 1917.

[16] Frank Natterer.
The Mathematics of Computerized Tomography.
Classics in Applied Mathematics. Wiley, 1986.

[17] Frank Natterer and Frank Wübbeling.
Mathematical Methods in Image Reconstruction.
SIAM, 2001.

[18] Gabor T. Herman.
Image Reconstruction from Projections: The Fundamentals of Computerized Tomography.
Academic Press, 1980.

[19] Avinash C. Kak and Malcolm Slaney.
Principles of Computerized Tomographic Imaging.
IEEE Press, 1988.

[20] Paolo Facchi, Marilena Ligabò, and Sergio Solimini.
Tomography: mathematical aspects and applications.
Physica Scripta, 90(7):074007, June 2015.

[21] Richard C. Aster, Brian Borchers, and Clifford H. Thurber.
Parameter Estimation and Inverse Problems.
Elsevier, 2nd edition, 2011.

[22] Mario Bertero and Patrizia Boccacci.
Introduction to Inverse Problems in Imaging.
IOP Publishing, 1998.

[23] Ralf Hielscher, Daniel Potts, Jürgen Prestin, Helmut Schaeben, and Matthias Schmalz.
The radon transform on so(3): A fourier slice theorem and numerical inversion.
Inverse Problems, 24(2):025011, 2008.

[24] Eric Todd Quinto.
An introduction to x-ray tomography and radon transforms.
In *Proceedings of Symposia in Applied Mathematics*, pages 1–45, 2006.

[25] S.R. Deans.
The Radon Transform and Some of Its Applications.
Dover Books on Mathematics Series. Dover Publications, Incorporated, 2013.

[26] W. Stiller.
Basics of iterative reconstruction methods in computed tomography.
European Journal of Radiology, 102:196–205, 2018.

[27] M. J. Willemink and et al.
The evolution of image reconstruction for ct – from filtered back projection to artificial intelligence.
European Radiology, 29:2185–2195, 2019.

[28] Maria Lyra and Agapi Ploussi.
Filtering in spect image reconstruction.
International Journal of Biomedical Imaging, 2011(1):693795.

[29] Marcel Beister, Daniel Kolditz, and Willi A. Kalender.
Iterative reconstruction methods in x-ray ct.
Physica Medica: European Journal of Medical Physics, 28(2):94–108, Apr 2012.

[30] F. Usman, R. Zainon, A. Saidu, and A. Bala.
Evaluation of image reconstruction techniques in single photon emission computed tomography (spect/ct) imaging.
Research & Reviews: Journal of Physics, 5(2):13–22, 2016.

[31] F. H. P. van Velden and et al.
A study of lesion contrast recovery for iterative pet image reconstructions versus filtered backprojection using an anthropomorphic thoracic phantom.
Physics in Medicine and Biology, 47(21):3809–3824, 2002.

[32] Alain Seret and Julien Forthomme.
Comparison of different types of commercial filtered backprojection and ordered-subset expectation maximization spect reconstruction software.
Journal of Nuclear Medicine Technology, 37(3):179–187, 2009.

[33] H. W. A. M. de Jong and M. Lubberink.
Evaluation of iterative reconstruction (osem) versus filtered back-projection for the assessment of myocardial glucose uptake and perfusion using dynamic pet.
European Journal of Nuclear Medicine and Molecular Imaging, 34(3):316–319, 2007.

[34] H. M. Hudson and R. S. Larkin.
Accelerated image reconstruction using ordered subsets of projection data.
IEEE Transactions on Medical Imaging, 13(4):601–609, 1994.

[35] Jeroen Bertels, David Robben, Robin Lemmens, and Dirk Vandermeulen.
Convolutional neural networks for medical image segmentation, 2022.

[36] Suraj Pai, Ibrahim Hadzic, Dennis Bontempi, Keno Bressem, Benjamin H. Kann, Andriy Fedorov, Raymond H. Mak, and Hugo J. W. L. Aerts.
Vision foundation models for computed tomography, 2025.

[37] DR Sarvamangala and Raghavendra V Kulkarni.
Convolutional neural networks in medical image understanding: a survey.
Evolutionary intelligence, 15(1):1–22, 2022.

[38] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille.
Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs.
arXiv preprint arXiv:1606.00915, 2016.

[39] Olaf Ronneberger, Philipp Fischer, and Thomas Brox.
U-net: Convolutional networks for biomedical image segmentation.
arXiv preprint arXiv:1505.04597, 2015.

[40] Md Zahangir Alom, Mahmudul Hasan, Chris Yakopcic, Tarek M. Taha, and Vijayan K. Asari.
Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation.
arXiv preprint arXiv:1802.06955, 2018.

[41] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik.
Rich feature hierarchies for accurate object detection and semantic segmentation.
In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 580–587, 2014.

[42] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun.
Faster r-cnn: Towards real-time object detection with region proposal networks.
In *Advances in Neural Information Processing Systems*, pages 91–99, 2015.

[43] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick.
Mask r-cnn.
arXiv preprint arXiv:1703.06870, 2017.

[44] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi.
You only look once: Unified, real-time object detection.
In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[45] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yu Fu, and Alexander C. Berg.
Ssd: Single shot multibox detector.
In *European Conference on Computer Vision*, pages 21–37, 2016.

[46] Jonathan Masci, Ueli Meier, Dan Cireşan, and Jürgen Schmidhuber.
Stacked convolutional auto-encoders for hierarchical feature extraction.
International Conference on Artificial Neural Networks, 2011.

[47] Jonathan Long, Evan Shelhamer, and Trevor Darrell.
Fully convolutional networks for semantic segmentation.
In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015.

[48] Bo Zhu, Jeremiah Z. Liu, Stephen F. Cauley, Bruce R. Rosen, and Matthew S. Rosen.
Image reconstruction by domain-transform manifold learning.
Nature, 555(7697):487–492, 2018.

[49] Bingyu Zhu, Jeremiah Z. Liu, Matthew S. Rosen, and Bruce R. Rosen.
iradonmap: A deep learning framework for radon inversion.
IEEE Transactions on Medical Imaging, 38(6):1336–1348, 2019.

[50] Pierre Baldi.
Autoencoders, unsupervised learning, and deep architectures.
In *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, pages 37–49. 2012.

[51] Guansong Pang, Chunhua Shen, Longbing Cao, and Anton van den Hengel.
A comprehensive survey of autoencoders and their applications.
IEEE Transactions on Knowledge and Data Engineering, 2021.

[52] Andrea Asperti, Davide Evangelista, and Elena Loli Piccolomini.
A survey on variational autoencoders from a green ai perspective.
SN Computer Science, 2(4):301, May 2021.

[53] Diederik P. Kingma and Max Welling.
Auto-encoding variational bayes.
International Conference on Learning Representations, 2014.

[54] Xi Chen, Diederik P. Kingma, Tim Salimans, Yan Duan, Prafulla Dhariwal, John Schulman, Ilya Sutskever, and Pieter Abbeel.
Variational lossy autoencoder.
International Conference on Learning Representations, 2017.

[55] Cong Ma, Chao Li, and Yu Zhang.
Medical image compression using variational autoencoders.
Multimedia Tools and Applications, 2020.

[56] Sovanlal Mukherjee, Anurima Patra, Hala Khasawneh, Panagiotis Korfiatis, Naveen Rajamohan, Garima Suman, Shounak Majumder, Ananya Panda, Matthew P. Johnson, Nicholas B. Larson, Darryl E. Wright, Timothy L. Kline, Joel G. Fletcher, Suresh T. Chari, and Ajit H. Goenka.
Radiomics-based machine-learning models can detect pancreatic cancer on prediagnostic computed tomography scans at a substantial lead time before clinical diagnosis.
Gastroenterology, 163(5):1435–1446.e3, 2022.

[57] Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger.
3d u-net: Learning dense volumetric segmentation from sparse annotation.
In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 424–432, 2016.

[58] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang.
Unet++: A nested u-net architecture for medical image segmentation.

IEEE Transactions on Medical Imaging, 39(6):1856–1867, 2020.

[59] Fabian Isensee, Paul F. Jaeger, Simon A. A. Kohl, Jens Petersen, and Klaus H. Maier-Hein.
nnu-net: A self-configuring method for deep learning-based biomedical image segmentation.
Nature Methods, 18:203–211, 2021.

[60] Anil B. Gavade, Rajendra Nerli, Neel Kanwal, Priyanka A. Gavade, Shridhar Sunilkumar Pol, and Syed Tahir Hussain Rizvi.
Automated diagnosis of prostate cancer using mpmri images: A deep learning approach for clinical decision support.
Computers, 12(8), 2023.

[61] Angelo Didonna, Dayron Ramos Lopez, Giuseppe Iaselli, Nicola Amoroso, Nicola Ferrara, and Gabriella Maria Incoronata Pugliese.
Deep convolutional framelets for dose reconstruction in boron neutron capture therapy with compton camera detector.
Cancers, 17(1), 2025.

[62] Eunhee Kang, Won Chang, Jaejun Yoo, and Jong Chul Ye.
Deep convolutional framelet denosing for low-dose ct via wavelet residual network.
IEEE Transactions on Medical Imaging, 37(6):1358–1369, 2018.

[63] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin.
Attention is all you need.
In *Advances in Neural Information Processing Systems*, 2017.

[64] Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Hervé Jégou.
Training data-efficient image transformers & distillation through attention.
In *International Conference on Machine Learning*, 2021.

[65] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo.
Swin transformer: Hierarchical vision transformer using shifted windows.
In *International Conference on Computer Vision*, 2021.

[66] Jieneng Chen, Yufan Lu, Qihang Yu, Xinzhi Luo, Ehsan Adeli, Yutong Wang, Le Lu, Alan L. Yuille, and Yuyin Zhou.
Transunet: Transformers make strong encoders for medical image segmentation.
In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2021.

[67] Ali Hatamizadeh, Yucheng Tang, Vishwesh Nath, Dong Yang, Andriy Myronenko, Bennett Landman, Holger Roth, and Daguang Xu.
Unetr: Transformers for 3d medical image segmentation.
In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[68] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby.
An image is worth 16x16 words: Transformers for image recognition at scale.
In *International Conference on Learning Representations*, 2021.

[69] Hangbo Bao, Li Dong, and Furu Wei.
Beit: Bert pre-training of image transformers.
In *International Conference on Learning Representations*, 2022.

[70] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang.
Restormer: Efficient transformer for high-resolution image restoration.
In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[71] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick.
Masked autoencoders are scalable vision learners.
In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[72] Wenhai Wang, Enze Xie, Xiang Li, Deng-Ping Fan, Kaitao Song, Ding Liang, Tong Lu, Ping Luo, and Ling Shao.
Pyramid vision transformer: A versatile backbone for dense prediction without convolutions.
In *International Conference on Computer Vision*, 2021.

[73] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M. Alvarez, and Ping Luo.
Segformer: Simple and efficient design for semantic segmentation with transformers.
In *NeurIPS*, 2021.

[74] M. Boal, C. Giovene Di Girasole, F. Tesfai, T. E. M. Morrison, S. Higgs, J. Ahmad, A. Arezzo, and N. Francis.
Evaluation status of current and emerging minimally invasive robotic surgical platforms.
Surgical Endoscopy, 38(2):554–585, Feb 2024.

[75] Filippo Attivissimo, Attilio Di Nisio, Anna Maria Lucia Lanzolla, and Mattia Alessandro Ragolia.
Analysis of position estimation techniques in a surgical em tracking system.
IEEE Sensors Journal, 21(13):14389–14396, 2021.

[76] Gregorio Andria, Filippo Attivissimo, Attilio Di Nisio, Anna Maria Lucia Lanzolla, and Mattia Alessandro Ragolia.
Assessment of position repeatability error in an electromagnetic tracking system for surgical navigation.
Sensors, 20(4), 2020.

[77] Gregorio Andria, Filippo Attivissimo, Attilio Di Nisio, Anna M. L. Lanzolla, Giuseppe Guglielmi, and Rita Terlizzi.
Dose optimization in chest radiography: System and model characterization via experimental investigation.
IEEE Transactions on Instrumentation and Measurement, 63(5):1163–1170, 2014.

[78] Attilio Di Nisio, Nicola Giaquinto, Anna Maria Lucia Lanzolla, Mattia Alessandro Ragolia, Marco Scarpetta, and Sandro Carrara.
Platinum nanostructured needle-shaped sensors for ion detection in biomedical applications.
IEEE Sensors Journal, 22(23):22404–22412, 2022.

[79] Seenia Francis, PB Jayaraj, PN Pournami, and Niyas Puzhakkal.
Contourgan: Auto-contouring of organs at risk in abdomen computed tomography images using generative adversarial network.
International Journal of Imaging Systems and Technology, 33(5):1494–1504, 2023.

[80] Shangzhu Jin, Sheng Yu, Jun Peng, Hongyi Wang, and Yan Zhao.
A novel medical image segmentation approach by using multi-branch segmentation network based on local and global information synchronous learning.
Scientific Reports, 13(1):6762, 2023.

[81] Jonathan Long, Evan Shelhamer, and Trevor Darrell.
Fully convolutional networks for semantic segmentation.
In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3431–3440, 2015.

[82] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert.
Attention u-net: Learning where to look for the pancreas, 2018.

[83] Yuyin Zhou, Lingxi Xie, Elliot K. Fishman, and Alan L. Yuille.
Deep supervision for pancreatic cyst segmentation in abdominal ct scans.
In Maxime Descoteaux, Lena Maier-Hein, Alfred Franz, Pierre Jannin, D. Louis Collins, and Simon Duchesne, editors, *Medical Image Computing and Computer Assisted Intervention MICCAI 2017*, pages 222–230, Cham, 2017. Springer International Publishing.

[84] Zhihui Guo, Ling Zhang, Le Lu, Mohammadhadi Bagheri, Ronald M. Summers, Milan Sonka, and Jianhua Yao.
Deep logismos: Deep learning graph-based 3d segmentation of pancreatic tumors on ct scans.
In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1230–1233, 2018.

[85] Amber L Simpson, Michela Antonelli, Spyridon Bakas, Michel Bilello, Keyvan Farahani, Bram Van Ginneken, Annette Kopp-Schneider, Bennett A Landman, Geert Litjens, Bjoern Menze, et al.
A large annotated medical image dataset for the development and evaluation of segmentation algorithms.
arXiv preprint arXiv:1902.09063, 2019.

[86] https://doi.org/10.7937/k9/tcia.2018.sc20fo18.

[87] Yan Zhu, Peijun Hu, Xiang Li, Yu Tian, Xueli Bai, Tingbo Liang, and Jingsong Li.
An end-to-end data-adaptive pancreas segmentation system with an image quality control toolbox.
Journal of Healthcare Engineering, 2023(1):3617318, 2023.

[88] Yue Du, Xiaoying Zuo, Shidong Liu, Dai Cheng, Jie Li, Mingzhu Sun, Xin Zhao, Hui Ding, and Yabin Hu.
Segmentation of pancreatic tumors based on multi-scale convolution and channel attention mechanism in the encoder-decoder scheme.
Medical Physics, 50(12):7764–7778, 2023.

[89] Jie Xue, Kelei He, Dong Nie, Ehsan Adeli, Zhenshan Shi, Seong-Whan Lee, Yuanjie Zheng, Xiyu Liu, Dengwang Li, and Dinggang Shen.
Cascaded multitask 3-d fully convolutional networks for pancreas segmentation.
IEEE Transactions on Cybernetics, 51(4):2153–2165, 2021.

[90] Connor Shorten and Taghi M. Khoshgoftaar.
A survey on image data augmentation for deep learning.
Journal of Big Data, 6(1), 2019.

[91] Shachar Kaufman, Saharon Rosset, and Claudia Perlich.
Leakage in data mining: Formulation, detection, and avoidance.
In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2012.

[92] Sayash Kapoor and Arvind Narayanan.
Leakage and the reproducibility crisis in machine learning.
In *ACM Conference on Fairness, Accountability, and Transparency (FAccT)*, 2022.

[93] Kun Fang, Baochun He, Libo Liu, Haoyu Hu, Chihua Fang, Xuguang Huang, and Fucang Jia.
Umrformer-net: a three-dimensional u-shaped pancreas segmentation method based on a double-layer bridged transformer network.
Quantitative Imaging in Medicine and Surgery, 13(3):1619, 2023.

[94] Tongtong Zhao, Zhaonan Sun, Ying Guo, Yumeng Sun, Yaofeng Zhang, and Xiaoying Wang.
Automatic renal mass segmentation and classification on ct images based on 3d u-net and resnet algorithms.
Frontiers in Oncology, 13:1169922, 2023.

[95] Luyang Cao and Jianwei Li.
Strongly representative semantic-guided segmentation network for pancreatic and pancreatic tumors.

Biomedical Signal Processing and Control, 87:105562, 2024.

[96] Luyang Cao, Jianwei Li, and Shu Chen.
Multi-target segmentation of pancreas and pancreatic tumor based on fusion of attention mechanism.
Biomedical Signal Processing and Control, 79:104170, 2023.

[97] Yang Li, Bobo Yan, Jianxin Hou, Bingyang Bai, Xiaoyu Huang, Canfei Xu, and Limei Fang.
Unet based on dynamic convolution decomposition and triplet attention.
Scientific Reports, 14(1):271, Jan 2024.
https://github.com/shuaihou1234/DTA-UNet.

[98] Kai Cao, Yingda Xia, Jiawen Yao, Xu Han, Lukas Lambert, Tingting Zhang, Wei Tang, Gang Jin, Hui Jiang, Xu Fang, Isabella Nogues, Xuezhou Li, Wenchao Guo, Yu Wang, Wei Fang, Mingyan Qiu, Yang Hou, Tomas Kovarnik, Michal Vocka, Yimei Lu, Yingli Chen, Xin Chen, Zaiyi Liu, Jian Zhou, Chuanmiao Xie, Rong Zhang, Hong Lu, Gregory D. Hager, Alan L. Yuille, Le Lu, Chengwei Shao, Yu Shi, Qi Zhang, Tingbo Liang, Ling Zhang, and Jianping Lu.
Large-scale pancreatic cancer detection via non-contrast ct and deep learning.
Nature Medicine, 29(12):3033–3043, 2023.

[99] K. Laxminarayanamma, R. V. Krishnaiah, and P. Sammulal.
Idrcnn: A novel deep learning network model for pancreatic ductal adenocarcinoma detection on computed tomography.
International Journal of Electrical and Computer Engineering Systems, 15(1):1–11, 2024.

[100] Nicola Ferrara, Gregorio Andria, Marco Scarpetta, Anna Maria Lucia Lanzolla, Filippo Attivissimo, Attilio Di Nisio, and Dayron Ramos.
2d and 2.5d pancreas and tumor segmentation in heterogeneous ct images of pdac patients.
In *2024 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pages 1–5, 2024.

[101] Shizhan Gong, Yuan Zhong, Wenao Ma, Jinpeng Li, Zhao Wang, Jingyang Zhang, Pheng-Ann Heng, and Qi Dou.
3dsam-adapter: Holistic adaptation of sam from 2d to 3d for promptable medical image segmentation.

arXiv e-prints, pages arXiv–2306, 2023.

[102] P. Terrón, I. Perlado, N. Romero, F. Araujo, and J. Pascual.
Boron neutron capture therapy followed by image-guided intensity-modulated radiotherapy for locally recurrent head and neck cancer: A prospective phase i/ii trial.
Cancers, 15(10), 2025.

[103] Author(s) of Systematic Review.
Efficacy and safety of boron neutron capture therapy: A systematic review.
Radiation Research, 2025.

[104] N. Romero F. Araujo P. Terrón, I. Perlado and J. Pascual.
Optimized beam shaping assembly for a 2.1-mev proton-accelerator-based neutron source for boron neutron capture therapy.
Scientific Reports, 11, 2021.

[105] Fauzi Arief, Tsurayya Afifah Hana, Harish Ahmad Faisal, and Wijaya Gede Sutresna.
Beam shaping assembly optimization for boron neutron capture therapy facility based on cyclotron 30mev as neutron source.
ASEAN Journal on Science and Technology for Development, 35(3), 2018.

[106] Authors at University of Osaka.
Neutronics design of a beam shaping assembly in an accelerator-based neutron source for the boron neutron capture therapy system at the university of osaka.
Applied Sciences, 15(18):10248, 2025.

[107] OICC Press Authors.
Beam shaping assembly study for bnct facility based on a 2.5mev proton accelerator on li target.
Journal of Theoretical and Applied Physics, 12:249–256, 2018.

[108] Yongquan Wang, Zezhen Wang, Ning Li, Xingcai Guan, and Long Gu.
Design of beam shaping assembly for accelerator-based boron neutron capture therapy and study on its clinical parameter.
Journal of Atomic Energy Science and Technology, 56(7):1440–1447, 2022.

[109] Jason K. Rockhill and George E. Laramore.

Chapter 20 - neutron radiotherapy.
In Leonard L. Gunderson and Joel E. Tepper, editors, *Clinical Radiation Oncology (Third Edition)*, pages 377–384. W.B. Saunders, Philadelphia, third edition edition, 2012.

[110] Authors of Dongguan Neutron Science Center.
Design of beam shaping assembly based on 3.5 mev radio-frequency quadrupole proton accelerator for boron neutron capture therapy.
Acta Physica Sinica, 2018.

[111] Guangru Li, Wei Jiang, Lu Zhang, Weiqiang Chen, and Qiang Li.
Design of beam shaping assemblies for accelerator-based bnct with multi-terminals.
Frontiers in Public Health, 2021.

[112] Anita Caracciolo, Luca Buonanno, Davide Di Vita, Ilenia D'Adda, Andrew Chacon, Marissa Kielly, Marco Carminati, Mitra Safavi-Naeini, and Carlo Fiorini.
Benedicte (boron neutron capture): A versatile gamma-ray detection module for boron neutron capture therapy.
IEEE Transactions on Radiation and Plasma Medical Sciences, 6(8):886–892, 2022.

[113] T. Ferri, A. Caracciolo, G. Borghi, M. Carminati, S. Altieri, N. Protti, and C. Fiorini.
Imaging study and first measurements of a labr3 gamma detector for bnct applications.
IEEE Transactions on Nuclear Science, 71(5):1019–1025, 2024.

[114] T. Ferri, A. Caracciolo, F. Ghisio, M. Piroddi, M. Pandocchi, C. Fiorini, M. Carminati, V. Pascali, N. Protti, D. Mazzucconi, L. Grisoni, D. Ramos, N. Ferrara, K. Thielemans, and G. Borghi.
Design and validation of a spect prototype for treatment monitoring in bnct and first experimental tomographic results.
IEEE Transactions on Radiation and Plasma Medical Sciences, 10(1):126–136, 2026.

[115] C. Ahdida, D. Bozzato, D. Calzolari, F. Cerutti, N. Charitonidis, A. Cimmino, A. Coronetti, G. L. D'Alessandro, A. Donadon Servelle, L. S. Esposito, R. Froeschl, R. García Alía, A. Gerbershagen, S. Gilardoni, D. Horváth, G. Hugo, A. Infantino, V. Kouskoura, A. Lechner, B. Lefebvre, G. Lerner, M. Magistris, A. Manousos, G. Moryc, F. Ogallar Ruiz, F. Pozzi, D. Prelipcean, S. Roesler,

R. Rossi, M. Sabaté Gilarte, F. Salvat Pujol, P. Schoofs, V. Stránský, C. Theis, A. Tsinganis, R. Versaci, V. Vlachoudis, A. Waets, and M. Widorski.
New capabilities of the fluka multi-purpose code.
Frontiers in Physics, Volume 9 - 2021, 2022.

[116] J. Allison, K. Amako, J. Apostolakis, P. Arce, M. Asai, T. Aso, E. Bagli, A. Bagulya, S. Banerjee, G. Barrand, B.R. Beck, A.G. Bogdanov, D. Brandt, J.M.C. Brown, H. Burkhardt, Ph. Canal, D. Cano-Ott, S. Chauvie, K. Cho, G.A.P. Cirrone, G. Cooperman, M.A. Cortés-Giraldo, G. Cosmo, G. Cuttone, G. Depaola, L. Desorgher, X. Dong, A. Dotti, V.D. Elvira, G. Folger, Z. Francis, A. Galoyan, L. Garnier, M. Gayer, K.L. Genser, V.M. Grichine, S. Guatelli, P. Guèye, P. Gumplinger, A.S. Howard, I. Hřivnáčová, S. Hwang, S. Incerti, A. Ivanchenko, V.N. Ivanchenko, F.W. Jones, S.Y. Jun, P. Kaitaniemi, N. Karakatsanis, M. Karamitros, M. Kelsey, A. Kimura, T. Koi, H. Kurashige, A. Lechner, S.B. Lee, F. Longo, M. Maire, D. Mancusi, A. Mantero, E. Mendoza, B. Morgan, K. Murakami, T. Nikitina, L. Pandola, P. Paprocki, J. Perl, I. Petrović, M.G. Pia, W. Pokorski, J.M. Quesada, M. Raine, M.A. Reis, A. Ribon, A. Ristić Fira, F. Romano, G. Russo, G. Santin, T. Sasaki, D. Sawkey, J.I. Shin, I.I. Strakovsky, A. Taborda, S. Tanaka, B. Tomé, T. Toshito, H.N. Tran, P.R. Truscott, L. Urban, V. Uzhinsky, J.M. Verbeke, M. Verderi, B.L. Wendt, H. Wenzel, D.H. Wright, D.M. Wright, T. Yamashita, J. Yarba, and H. Yoshida.
Recent developments in geant4.
Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 835:186–225, 2016.

[117] L. Santini, T. Ferri, L. Grisoni, A. Caracciolo, D. Bortot, G. Borghi, A. Pola, S. Agosteo, V. Pascali, N. Protti, D.Ramos Lòpez, N. Ferrara, and D. Mazzucconi.
Development and simulation of a spect real time dose monitoring system for bnct: response at the lena reactor.
Physica Medica, 137:105070, 2025.

[118] T. Ferri, F. Rosellini, A. Caracciolo, G. Borghi, M. Carminati, F. Camera, A. Giaz, and C. Fiorini.
Gamma-ray position-of-interaction estimation in a thick monolithic labr3 detector using artificial neural networks.
IEEE Transactions on Radiation and Plasma Medical Sciences, 9(3):284–295, 2025.

[119] Martin Richard Jaekel, M. Capeans, I. Efthymiopoulos, A. Fabich, R. Guida, G. Maire, M. Moll, D. Pfeiffer, F. Ravotti, and H. Reithler.
CERN-GIF++: a new irradiation facility to test large-area particle detectors for the high-luminosity LHC program.
PoS, TIPP2014:102, 2015.

[120] Dorothea Pfeiffer, Georgi Gorine, Hans Reithler, Bartolomej Biskup, Alasdair Day, Adrian Fabich, Joffrey Germa, Roberto Guida, Martin Jaekel, and Federico Ravotti.
The radiation field in the gamma irradiation facility gif++ at cern.
Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 866:91–103, 2017.

[121] Rachel Margraf and Nikolaos Charitonidis.
Muon beam studies in the h4 beam line and the gamma irradiation facility (gif++).
2018.

[122] Dipanwita Banerjee, Johannes Bernhard, Markus Brugger, Nikolaos Charitonidis, Niels Doble, Lau Gatignon, and Alexander Gerbershagen.
The north experimental area at the cern super proton synchrotron.
2021.
Dedicated to Giorgio Brianti on the 50th anniversary of his founding the SPS Experimental Areas Group of CERN-Lab II and hence initiating the present Enterprise.

[123] S. Agosteo, S. Altieri, G. Belli, A. Bonifas, V. Carabelli, L. Gatignon, N. Hessey, M. Maggi, J. P. Peigneux, H. Reithler, M. Silari, P. Vitulo, and M. Wegner.
A facility for the test of large-area muon chambers at high rates.
Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 452(1):94–104, 2000.

[124] Automation und Messtechnik GmbH.
Datasheet for gamma probes 6150ad-15, 2025.

[125] Raafat Abdul H. Muslim, Saad Nafea Yaqoob, and Rawnq Qays Ghdban.
Refined formulation of the taylor equation for calculating build-up factors of gamma rays in aluminium and anticipated radiation effects.
Radiation Physics and Chemistry, 224:112089, 2024.

[126] Dariush Sardari, Samaneh Baradaran, Farshid Babapour Mofrad, and Nazanin Marzban.
Monte carlo calculation of buildup factors for 50 kev–15 mev photons in tungsten up to 15 mean free paths.
Applied Radiation and Isotopes, 183:110150, 2022.

[127] Omar Chibani.
New photon exposure buildup factors.
Nuclear Science and Engineering, 137(2):215–225, 2001.

[128] Radprocalculator.

List of Tables

List of Figures

Figure **Page**

Acknowledgements

This work has been supported by the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.4 - Call for tender No. 3138 of 2021/12/16 of Italian Ministry of Universities and Research funded by the European Union - NextGenerationEU Award Number: CN000013, Concession Decree No. 1031 of 2022/06/17 adopted by the Italian Ministry of Universities and Research, D93C22000430001, National Centre for HPC, Big Data and Quantum Computing.