POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

ANNs for 2D positioning and improved time resolution in monolithic TOF-PET detectors

Tesi di Laurea Magistrale in
Biomedical Engineering - Ingegneria Biomedica

Riccardo Pallone, 947244

Advisor:
Prof. Carlo Ettore Fiorini

Co-advisors:
Ilenia D'Adda

Academic year:
2021-2022

Abstract: Monolithic detectors for PET and TOF-PET imaging systems are gaining more and more interest due to their higher sensitivity, timing resolution, and spatial resolution. Moreover, the spatial propagation of the scintillation light can be exploited to gain additional information on the temporal dynamics of the detection. This current thesis work aims to provide a complete data processing algorithm that provides predictions on 2D gamma-interaction positioning and improves the detector time resolution. Two artificial neural networks (ANNs) are implemented: a positioning ANN predicts the scintillation position, and a timing ANN predicts a correction for time of flight (TOF) measurements. The geometry of the detector under consideration consists of a monolithic crystal of dimensions 50.8x50.8x20 mm^3 coupled to an 8x8 matrix of SiPMs. The simulation toolkit ANTS2 is used to collect data from different detector configurations, using $LaBr_3$(Ce) and LYSO(Ce) crystals together with different coatings. The study includes comparisons between different spatially-resolved training datasets, collected by moving a pencil beam source with 1 mm and 5 mm step sizes. Both positioning and timing data are analyzed to evaluate the best data preparation. The positioning ANN showed a 1.805 mm spatial FWHM on a LYSO(Ce) configuration. The correction predicted by the timing ANN was able to reduce the TOF uncertainty from the initial value of 467 ps FWHM to the corrected one of 192 ps FWHM for a considered $LaBr_3$(Ce) configuration.

Key-words: Positron emission tomography, TOF-PET, monolithic scintillator, machine learning, artificial neural network (ANN)

1. Introduction

Biomedical applications have always been a field in which multiple innovative engineering methods have succeeded in improving the quality of medical treatments. In particular, Medical Imaging and Artificial Neural Networks (ANN) are two research fields that have found interesting ways to improve each other, adapting each other to arrive at promising results over the past 20 years. In the wake of these successes, this research aims to make a contribution to the application of ANN to improve Positron Emission Tomography (PET) detector performances.
This first chapter contains a brief summary of PET and TOF-PET functioning principles, together with a description of basic PET detector's components. An introduction on Machine Learning methods follows, along

with some applications of Machine Learning in PET. Finally a review of ANNs implemented for PET describes the state of the art of this topic.

1.1. Introduction to PET and gamma detectors

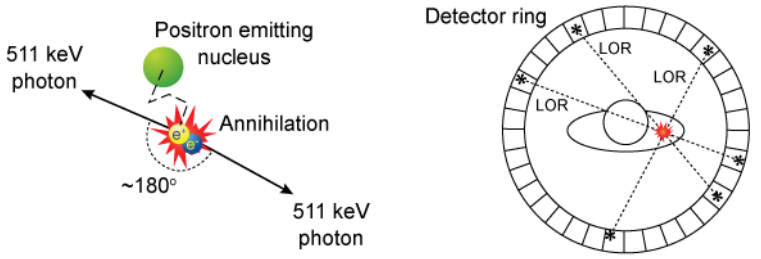
Positron Emission Tomography (PET) is one of the most common molecular imaging techniques. It is used to reconstruct a three-dimensional image of the metabolic activity of the body by using a radioactive biomarker that follows a specific metabolic pathway. The first PET scanner for human application goes back to 1974. Afterward, there was a growth in the production of new radiopharmaceuticals for use as PET radioactive tracers. Further advancements in this technique resulted in valuable predictions and treatments of several types of cancers, tumors, strokes, epilepsies, and Alzheimer's disease [1].
Clinical PET devices are often coupled with Computed Tomography (CT) and Magnetic Resonance Imaging (MRI), in order to obtain an enriched image from the patient.

1.1.1 PET principles

Positron emission tomography is a medical imaging modality that measures the distribution of a positron-emitting isotope inside a patient body. The patient is first injected with a tracer containing a positron-emitting radionuclide (such as ^{11}C and ^{18}F), the purpose of the tracer is to bind with the molecules of interest e.g. receptors on tumoral cells. After the tracer is introduced, the subject is positioned into the PET machine, which consists of a cylindrical detector, composed of many adjacent rings, placed around a table where the patient is lied down.
The radionuclide in the tracer decays by converting a proton into a neutron, this process is associated with the emission of a positron. The positron travels a short distance (of about ~1.5 mm and called *positron range*) before annihilating with a nearby electron, each annihilation produces two 511 keV photons traveling in opposite directions (Figure 1a). This type of decay is called β^+ decay: a proton p of the radioisotope X decays in a neutron n, a positron e^+ and an electronic neutrino υ_e [2].

$$ {}^{A}_{Z}X_N \xrightarrow{\tau_A} {}^{A}_{Z-1}Y_{N+1} + e^+ + \nu_e \tag{1} $$

Subsequently, the two gamma rays interact with two opposite detectors in the PET machine, and a certain coincidence time window (2-6 ns depending on the detector geometry) is chosen to classify two detected photons as produced by the same annihilation. Finally, it is assumed that the annihilation takes place somewhere along the line between the two detectors, which is referred to as line of response (LOR) (Figure 1b) [3].
Two sources of uncertainty in position estimation come from the physics of the process: the positron range sets a lower limit to PET spatial resolution, additionally, variations in the momentum of the interacting particles result in an angular uncertainty in the direction of the 511 keV photons of around 4 mrad [4].



(a) *Annihilation process.* (b) *Tomographic reconstruction using LORs.*

Figure 1: *Imaging principle of PET [5].*

Errors can occur when two photons appear to be coincident, but they are not generated by the same annihilation event, this can be reduced by shortening the coincidence time window if the detector's time resolution allows it. Another undesired event is associated to scattered photons. Scattering happens when a photon originated from annihilation interacts with matter by means of Compton scattering. During this process, the energy of the 511 keV photon decreases according to the Klein–Nishina formula [3]:

$$ \frac{E^{'}}{E_\gamma} = \frac{1}{2 - \cos\theta} \tag{2} $$

where E_γ is the energy of the incident photon, $E^{'}$ is the photon energy after scattering and θ is the angle of deviation. Modern PET scanners have a sufficient energy resolution (around 10%) to reject part of these scattered photons. A typical PET scanner has a 1:1:1 ratio between true coincidences, scattered events, and false coincidences.

1.1.2 PET system components: scintillators

The majority of detectors used in PET systems are based on scintillator crystals coupled with photodetectors. Scintillators are high-density crystals able to absorb gamma photons and emit scintillation photons at lower energy, which are in turn detected by photodetectors, an electric signal is produced and is elaborated by a readout circuit [6].
Scintillators and photodetectors in gamma cameras (detectors for gamma rays) can be implemented mainly in two different ways:

- *Segmented scintillator*: several blocks of scintillation material are placed after the collimator to detect the incoming radiation. Each block is optically isolated from from the others, and it can emit photons only on a single photodetector; in this way, the number of photodetectors used is equal to the number of pixels forming the image. This quantity is strictly correlated to the camera's resolution which became a fixed value depending on the number of photodetectors.

- *Monolithic scintillator*: it's the case of the Anger camera where a single block of scintillation material is used. Photodetectors are placed just after the scintillator, and they can be wider and fewer than the ones used in segmented scintillators. Using this disposition, an interaction event in the scintillator can be detected by more than one photodetector, but each of them responds with different intensity based on the relative position with the interaction event. This information is used by an algorithm in order to extrapolate the position of the interaction.

Another implementation exists, the *optical multiplexing*, which is a hybrid of the previous two, and it consists of multiple crystals that scintillate on a light guide, which in turn spreads the light over multiple photodetectors [3].
To evaluate the use of a certain scintillator in PET some properties must respect specific values to achieve an optimal conversion. A linear conversion is preferred, a fast response is needed, and good optical quality is required.
The process of prompt fluorescence is the key characteristic of the scintillation process, it is the emission of a visible or UV photon from an atom following its excitation. A similar process is phosphorescence which emits a longer wavelength light in a longer time. Delayed fluorescence, instead, presents a much longer emission time. To have an optimal use for PET systems, a scintillator should prefer the prompt fluorescence as a mechanism to emit light in order to generate fast signal pulses. Interactions with phosphorescence or delayed fluorescence mechanisms produce long-lived light which cannot be distinguished from noise [7].
Two main types of scintillators can be used:

- *Organic* scintillators are based on organic molecules with the peculiarity of forming a π-electron structure because of certain symmetry properties. In a condition of room temperature, all electrons are in the singlet S_{00} state. After the interaction with a gamma-ray (or other ionizing radiation in general) electrons are promoted to higher energy levels. Excited electrons above S_{10} are quickly de-excited to the S_{10} level through radiationless thermal conversion. From this state electrons mostly decay in S_0 levels through prompt fluorescence. Some excited singlet states may be converted into triplet states T_1, from which electrons can decay through phosphorescence. Because T_1 is under S_{10}, the emitted photons have lower energy, hence a longer wavelength. While in T_1 some molecules may be thermally excited back to S_1 state and then decay through normal fluorescence; this is the origin of the delayed fluorescence. Organic scintillators are transparent to their own fluorescence emission because the energy required for excitation is higher than most of the energy emitted with fluorescence; this characteristic ensures the arrival of emitted visible photons to photodetectors used in gamma cameras.
- The scintillation mechanism in *inorganic* crystals is based on energy states determined by the crystal lattice and the presence of impurities called activators. In the pure crystal, a valence and a conduction band are present. Electrons that decay from conduction to valence band would release photons of too high energy to lie in the visible range. The presence of activators inside the materials creates several energy levels in the gap that was first forbidden. Choosing a proper activator, a visible photon can be produced: when an electron-hole pair is produced from incoming radiation the positive hole will ionize the activator because the ionizing energy of impurities is lower than that of the crystal lattice, and a free electron can be captured by this ion creating a configuration with its own energy levels. These levels are closer

than the conductance-valence band, hence a visible photon can be emitted. Like organic scintillators, the emitted radiation, when activators are present, don't have enough energy to excite electrons in the valence band, hence the material is invisible to its own emission [8].

To evaluate the quality of a radiation detector such as the scintillator, some physical properties can be evalueted. The *Conversion efficiency* (CE) or *light yield Y* is the fraction of the absorbed energy actually converted in scintillation light. It is measured in the number of scintillation photons over the energy of the incoming radiation (usually ph/MeV). When dealing with high energy sources, like γ rays, we need to use a material capable of absorbing the radiation (511 KeV photons in PET) before it leaves the scintillator. *Density* ρ is related to the material's ability to interact with incoming photons through the Lambert-Beer Law: the higher is the density and higher the probability of absorption. Organic scintillators have usually a density close to that of water (from 0.8 to 1.2 gm/cm^3 on average), inorganic ones are denser (from 3 to 8 gm/cm^3 on average). Another crucial aspect of a scintillator is the *decay time* τ_{decay}. This parameter quantifies the average time for having a response after an interaction in the material. Both organics and inorganics scintillators can emit light in a range of the spectrum going from UV to infrared, according to the multiple ways of electrons de-excitation. Each material has one or more wavelengths called *peak emission wavelengths* for which the intensity is predominant. Given the properties mentioned above, inorganic scintillators are preferred for PET detectors given their high density and high effective atomic number Z_{eff}, that make them capable of stopping more gamma photons; their high light yield (up to 65 Kphe/MeV) helps the detector by increasing the signal to noise ratio. The drawback of inorganic scintillators is their longer response times compared to the responses of organic ones, this has to be taken into account when choosing a detector for TOF-PET.

Scintillator	ρ [g cm^{-3}]	Z_{eff}	τ_{decay} [ns]	Energy resolution (%FWHM)	Light yield [Kph/MeV]
BaF_2	4.9	54	0.8	8	9.4
$CeBr_3$	5.2	46	17	4	68
CsF	4.6	52	3	∼20	0.5
$LaBr_3$:Ce	5.1	45	16	3	63
L(Y)SO:Ce	7.1-7.4	65-66	38-44	8	29-32
L(Y)SO:Ce,Ca	7.1-7.4	65-66	31-39	8	34

Table 1: *Overview of scintillation crystals and they properties [9].*

1.1.3 PET system components: photodetectors

The wide use of scintillators in radiation detection would be impossible without the availability of photodetectors, devices aimed to convert the weak light signal coming from a scintillator into a corresponding electrical pulse.
A quite old and popular type of photodetector is the photomultiplier tube (PMTs), which is able to convert a few hundreds of visible photons into a current pulse composed of $\sim 10^7$ electrons. The incoming radiation hits a photocathode producing low-energy electrons. These photoelectrons are accelerated by an electrical field and multiplied by a series of structures called dynodes until they reach the anode. Finally, the electrical stimulus is collected and amplified with different methods following the purpose of the instrumentation.
Due to their compactness and low power consumption, solid-state photodetectors have become the most used detectors in PET acquisition systems, they provide a signal by collecting the charge liberated in the passage of the particle through a semiconductor [10]. In particular, Silicon Photomultipliers (SiPM) are receiving more and more attention for applications in the field of PET detectors, and they are the ones used in this research.
In order to illustrate the functioning of the SiPM, it is worth describing the principles behind the Avalanche Photodiodes (APD) and the Single-photon Avalanche Photodiodes (SPAD). An APD is a p-n photodiode biased in reverse direction below the breakdown voltage, i.e. the voltage beyond which the current in the diode increases exponentially, and the avalanche is self-sustained. When a photon hits the diode in the depletion region, it can generate an electron-hole pair. Then, these charge carriers are separated and accelerated by means of the high electric field applied, after reaching a sufficient velocity they can interact with the lattice to generate other electron-hole pairs. The repetition of this process is the so-called avalanche effect. SPADs consist of APDs biased above the breakdown voltage. The voltage applied is so high that any single incoming photon can trigger a self-sustained avalanche process. This operational configuration is called the Geiger mode, the signal generated is saturated by the multiplication stage, and must be restored by a quenching circuit in order to be able to detect new photons.
SiPMs are arrays of several SPADs (hundreds or thousands), that are connected in parallel and share a common output (Figure 2). Due to this configuration, the output of a SiPM is the sum of the current of each SPAD [11].

A single SPAD together with its quenching resistor form a microcell.

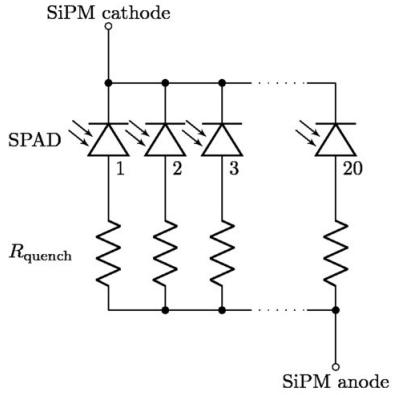


Figure 2: *Schematic example of a SiPM made up of 20 SPADs.*

Some figures of merit can be defined to understand the behavior of SiPMs and their suitability for PET systems.

- *Photon detection efficiency* (PDE) refers to the statistical probability that an incoming photon will trigger an avalanche process. It is expressed by the formula:

$$PDE = QE(\lambda) \times FF \times P_{triggering}(V) \tag{3}$$

  where the *quantum efficiency* (QE) expresses the probability that a photon impinging on the SiPM is actually transmitted to the silicon, absorbed in the silicon, and finally converted into an electron-hole pair. The QE is a function of the wavelength of the incoming photons. This dependency must be considered when coupling the SiPM with a particular scintillator with a specific wavelength emission, the combination of the two devices must guarantee the best transmission possible. The *fill factor* (FF) is the ratio of the sensitive part divided by the total area, it is less than unity as the physical separation of SPAD cells introduces considerable dead space. Finally, the *triggering probability* ($P_{triggering}$) is the probability for a photo-generated carrier of triggering an avalanche process, it is heavily dependent on the electrical field strength in the junction [12].
- *Gain* is defined as the total charge produced for each photon detected. It is calculated by dividing the charge established in the detector by the electron elementary charge [13]. A higher gain can be obtained by increasing the SPAD sizes and the over-voltage applied, i.e. the difference between the applied voltage and the breakdown value.
- *Recovery Time* is defined as the period of time until a cell is again fully sensitive after a breakdown.
- *Dark counts* refers to the signal produced in the total absence of light. It is possible for a charge carrier to be promoted from valence to conduction band and to generate an avalanche process; the main cause is thermal excitation, but tunneling effect can also happen. This signal is a source of noise that can be limited by lowering the operative temperature of the SiPM.
- *Crosstalk* occurs when one or more photons generated by the avalanche of a SPAD triggers a new avalanche in other ones. The direct crosstalk happens when the generated photons directly interact in the depletion region of adjacent SPADs, producing almost instantaneously secondary avalanches. Delayed crosstalk occurs when the generated photons produce an electron-hole pair outside the depleted region of the other SPAD, if the carries succeed in reaching the depleted region a secondary avalanche can be triggered with a random delay [14].

1.2. Introduction to TOF-PET principles

Time of flight (TOF) measurement was suggested as early as 1969 by Brownell et al. [15]. The first generation of TOF-PET scanners were developed in the 1980s when the primary application of PET was in brain and cardiac imaging using fast decaying isotopes [9]. TOF-PET scanners initially used crystals of CsF or BaF_2, but they both had low density that prevented the possibility of high spatial resolution and sampling. This problem led to the parallel development of non-TOF-PET during the late 70s and early 80s, which used BGO scintillators. The discovery of new scintillators in the 90s led to a new phase of TOF-PET research and development. Cerium-doped lutetium oxyorthosilicate or LSO (Lu_2SiO_5) has a sufficient high light yield (about 30000 photons/MeV) to obtain higher spatial resolution and a higher count rate and preserved high sensitivity.

This, combined with TOF information, pushed LSO to become the standard PET detector. The short decay time was immediately used to reduce the coincidence window from the 12 ns typical of BGO scanners to 6 ns, and it was later reduced to 4.5 ns with the development of faster electronics. The short decay time of LSO (40 ns) could be used for a new generation of TOF-PET scanners. In parallel, the more recent discovery of cerium-doped lanthanum bromide ($LaBr_3$) introduced new improvements due to its 16 ns decay time. A first commercial TOF-PET scanner, the Gemini TF PET/CT, was introduced by Philips in 2006 [16], and uses a LYSO scintillator crystal, which has a structure very similar to LSO. The time resolution of the Gemini TF is 585 ps if measured with a low activity source.

1.2.1 TOF-PET functioning

As explained in Section 1.1.1, the detection of a pair of back-to-back 511 keV photons in PET detectors determines an annihilation event to lie along the line of response (LOR). The exact location of the emission point along the LOR is given by the difference in detection times, or TOF difference, of the two annihilation photons $t_2 - t_1$ (Figure 3a). Since the TOF difference is proportional to the path length difference of the two photons, it can be used to know the exact position, along the LOR, of the positron annihilation (and if we ignore positron range, this is also the position of the radioisotope or tracer). The accuracy of this computation is obviously related to the time resolution Δt of the scanner. From simple cinematic, space and time uncertainty are proportional to each other by the relation $\Delta x \approx c\Delta t/2$. Time resolution is typically measured as the full-width-half-maximum (FWHM) of the time distribution of a point source [17], and it is usually reported as *coincidence rime resolution* (CTR) since it is usually derived from measurement with detectors placed in coincidence.
In conventional nonTOF-PET, the precision of TOF measurement (>1 ns) is not sufficient to provide a point along the LOR, the emission point has a uniform probability of lying anywhere along the LOR within the object. However, an accurate tomographic image of the radiotracer distribution can be reconstructed by using all LORs over the full azimuthal space together with an image reconstruction algorithm. The assumption of a uniform probability for the location of an emission point along the full LOR length will lead to noise in the final image since emissions from two different voxels will have overlapping LOR bins, hence they could be misidentified as a single event at the intersection of the two LORs (Figure 3b).

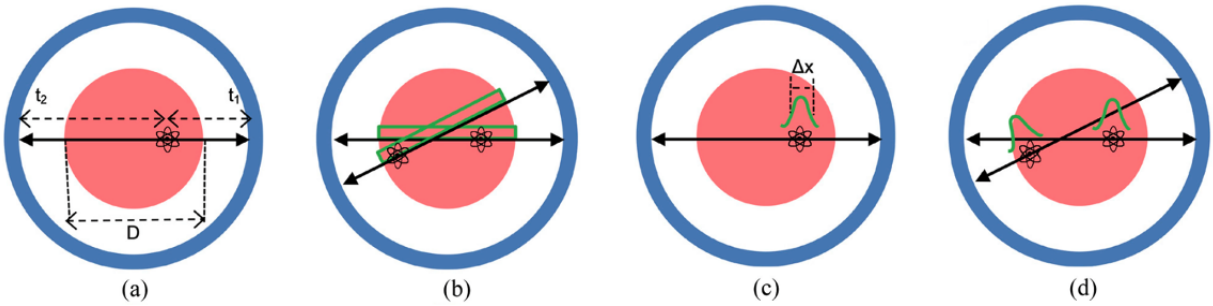


Figure 3: *(a) Annihilation point occurring within an object of diameter D. (b) In a non-TOF scanner, a uniform location probability along the LOR is assumed for the emission point, leading to overlapping LOr bins. (c) With improved TOF, the emission point is better localized along the LOR, with a precision that is defined by a Gaussian distribution of width Δx. (d) Improved localization of the emission points reduces (or avoids, as shown here) overlapping LOR bins. [9]*

Higher TOF resolutions allow more precise localization of the event along the LOR (Figure 3c), which leads to fewer overlapping LOR bins (Figure 3d), hence a higher signal-to-noise ratio (SNR) [9]. The first studies on TOF-PET demonstrated that TOF reduces the variance in the images of a factor proportional to $\Delta x/D$, where D is the size of the object to be imaged [18]. The improvement is called variance gain or TOF gain:

$$(\mathrm{SNR}_{\mathrm{gain}})^2 = \left(\frac{\mathrm{Var}_{\mathrm{nonTOF}}}{\mathrm{Var}_{\mathrm{TOF}}}\right) \cong \frac{D}{c/2\Delta t} = \frac{D}{\Delta x} \tag{4}$$

1.2.2 Timing Theory

The capability of improving tomographic images of TOF-PET detectors is primarily determined by the timing performance of the scintillation detector. Figure 4 describes some of the main factors that influence its time resolution.

The absorption of a gamma photon with energy E_γ at a certain time Θ results in the emission of a number of scintillation photons $N_e = E_\gamma Y$ (typically on the order of 10^4), with Y the light yeld of the scintillator (see Section 1.1.2). The temporal distribution $p_{t_d}(t \mid \Theta)$ of the N_d detected photons is given by the convolution of three probability density functions:

1. The shape of the scintillation light pulse (often characterized by exponential rise and decay time constants, τ_{rise} and τ_{decay}, respectively).
2. The *optical transfer time spread* (OTTS), which results from the transport of the scintillation photons within the crystal.
3. The *single-photon time resolution* (SPTR) of the photosensor, which determines the uncertainty with which the arrival time of a single photon at the sensor can be measured.

Other factors that may affect the time resolution include crosstalk, dark counts, noise, and the bandwidth and transient response of the readout electronics. From these considerations it can be deduced that the most appropriate scintillators for use in TOF-PET detectors are those that have a short rise and decay time constants to gain a rapid response, and high yield to increase SNR. To minimize the influence of OTTS, the scintillator should have high density ρ and high effective atomic number Z_{eff} in order to reduce the average path length of the annihilation quanta within the crystal until full absorption [19].
The readout of scintillators in TOF-PET detectors requires photosensors with internal gain (see Section 1.1.3) capable of detecting single photons with high PDE and SPTR. As mentioned in Section 1.1.3, the most recent TOF-PET scanners of essentially all commercial manufacturers are equipped with SiPMs because they allow to approach a time resolution of 200 ps FWHM, which is largely due to the excellent PDE and SPTR of SiPMs. The SPTR of a SiPM is determined primarily by the SPTR of its SPADs. State-of-the-art SiPMs have SPTR values in the range of 50–150 ps FWHM [20]. Regarding the output signal, to maintain a favorable pulse shape, the readout electronics must have sufficient bandwidth, as well as the lowest possible input impedance at signal frequencies [21].

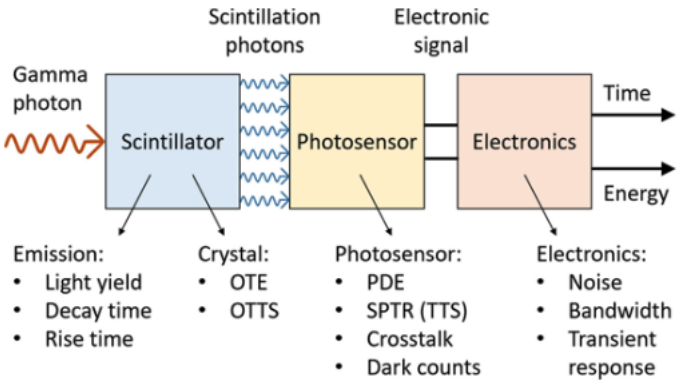


Figure 4: *Schematic representation of some of the main factors affecting detection time resolution. [9]*

1.3. Introduction to Machine Learning

Machine Learning (ML) is a field of data science that focuses on developing algorithms that best represent a set of data. ML uses subsets of data to generate an algorithm that may use novel or different combinations of features and weights than can be derived from the initial dataset. In ML, there are four different types of learning methods: supervised, unsupervised, semi-supervised, and reinforcement learning. The use of one or another depends on the data available and the type of problem that needs to be solved.

- **Supervised Learning**
  Each data is associated with a target, which is the feature to be predicted. Supervised learning uses patterns in the dataset to map features to the target so that an algorithm can make predictions on future datasets; this learning method is called supervised because the model computes an algorithm from feature-target pairs and checks using the target provided if the prediction is accurate or not. The dataset is initially split into training, validation, and testing datasets. Then the learning phase uses the training dataset combined with the validation datasets to tune the algorithm, by comparing performances on both datasets at each iteration. Finally, the test dataset is used to assess how the model behaves with unseen data.
  Supervised learning is suitable for regression and classification problems. The first one consists in predicting continuous numeric data, the second one wants to associate data with a set of categories.
- **Unsupervised learning**

This method does not use a target, it aims to detect and recognize patterns correlated with features within the dataset in order to categorize instances in the dataset. Using an unsupervised learning method can lead to discovering new features of the data.
Unsupervised learning problems are grouped into clustering and association problems. A clustering algorithm divides the dataset into clusters based on the features of the data inside it. There are several algorithms to perform the dataset division, like Hierarchical clustering, K-NN (k nearest neighbors), and Principal Component Analysis. Differently, association rules allow establishing associations amongst instances inside large databases.

- **Semisupervised learning**
  Semisupervised learning is useful with datasets that are composed of both labeled and unlabeled data. The labeled data are used to train a model, which is used on the unlabeled data to obtain a complete labeled dataset. The last step is to use the resulted labeled dataset to train a new model that performs better than the previous one.
- **Reinforcement Learning**
  Reinforcement learning is a technique able to learn from trial and error rather than data alone. The algorithm attempts different ways to solve a problem and it is "rewarded" when the output is close to the desired task.

1.4. Machine learning for PET

This research focuses on monolithic scintillators. These have some advantages over the segmented ones, where spatial resolution (that coincides with the crystal size) is limited by the size of the photodetector. The drawback for monolithic scintillators is that the light collection is not as efficient as that achieved by the 1:1 coupling between crystal and photodetector of segmented scintillators. This is one of the reasons why these detectors typically require robust and optimized algorithms capable of mapping scintillation coordinates with the distribution of light on the photodetectors.
Many positioning algorithms are applicable for developing 2D and 3D event positioning in a monolithic PET detector. The algorithms can be categorized into two types: theoretical methods and machine learning methods [22]. The disadvantage of the theoretical methods (e.g. non-linear analytical models) is their complicated calculations which might not be suitable to provide real-time estimation [23]. On the other hand, machine learning algorithms such as k-nearest neighbor (k-NN), gradient tree boosting (GTB) or ANNs are getting more and more interesting. Concerning TOF-PET detectors, temporal information is usually directly derived by the SiPM's readout electronic [24], however, machine learning algorithms can be implemented to reduce the temporal uncertainty that characterizes the detection of the scintillation photon (see Section 1.2.2). Training a machine learning algorithm for PET requires the acquisition of events from a 511-keV source with known beam geometry [25]. The relationship between the position and time of interaction and the spatial and temporal distribution of light is typically obtained using a dedicated calibration data acquisition that uses a pencil beam or fan-beam source of 511-keV photons to control the interaction location in the scintillator. For 3D position estimations, additional oblique or side-on irradiation is generally required. The calibration procedure must be repeated when changing detector configuration, such as crystal type and dimension, or SiPMs configuration.
This research focuses on ANNs for a monolithic detector, nevertheless, it is useful to provide also some examples of other machine learning methods used to improve positioning and timing in PET; only research dealing with detectors with monolithic scintillators will be reviewed, that is the same implementation of this research.
The main challenge in monolithic detectors is solving the inverse problem that maps the light distribution to the position-of-interaction in the presence of limited-statistics noise and the highly nonlinear behavior near the edges of the crystal, a problem well suited for machine learning methods [3]. A gradient tree boosting algorithm [26] has been used for 3D positioning in monolithic detectors in [27]. The choice of this method was driven by the operation of this algorithm: GBT algorithms rely only on binary decision operations, making them a relatively straightforward and a computationally efficient algorithm for fast event processing and satisfying the memory restrictions of FPGAs; for a 12 mm high LYSO-block, they achieved a spatial resolution of 1.4 mm FWHM, the study of F. Muller et al. also reported the limitation of GTB where no significant improvements on the positioning performance could be observed when the training data is greater than 5000 events per position. Several modified k-NN methods were tested in [28] and compared with the standard approach. The k-NN approaches were tested on a dataset measured with a 10 mm high LYSO detector, coupled with a SiPM array. It appeared that, depending on the number of reference events, 10% to 25% better spatial resolution can be obtained compared to the standard approach; in addition, improved methods yield the same spatial resolution as the standard method using 200 times less reference data, greatly reducing the time needed for both calibration and computation.

1.5. State of art of ANNs for PET

Different studies have pointed out the ability of ANN to effectively associate distributions with each other since it is a universal function approximator [29]. Not only the capability of intrinsically performing 3D positioning greatly increases the usability of an ANN but also the ability to learn more complicated relations (such as the one between signal intensity and temporal information) allows ANN to have the potential to outperform other machine learning algorithms. This is the reason why there is a lot of research on ANN applied to PET in the literature, however, it is noteworthy that while for position estimation there are multiple research with rather heterogeneous methods among them, for temporal estimation the number is less, and most of them try to link the profile of the response signal to the temporal information.
Table 2 illustrates results from researches on position estimation with ANN for PET.

	Crystal size [mm^3]	**Number of nodes in hidden layers**[2]	**Average x,y spatial FWHM [mm]**
F. Hashimoto et al. (2019)[1][30]	40×40×10	256-256-256	1.59
P. Conde et al. (2016) [31]	32.6×32.6×10	9-9	1.7
Y. Wang et al. (2013) [32]	25.5×25.5×10	8-8	1.86
P. Garcia de Acilu et al. (2012)[1][33]	18.5×21.4×10	4-4	2
M. Wedrowski (2010) [34]	20×20×10	5-5	2
P. Bruyndonckx et al. (2008) [35]	20×10×10	5-5	1.6
P. Bruyndonckx et al. (2004) [36]	20×10×10	5-5	1.9

[1] Data were collected by using simulation toolkits.
[2] Each number refers to one layer, input and output layers are not included.

Table 2: *Comparison of research on ANNs for monolithic PET detectors positioning [22].*

The performance of the reported ANNs is evaluated as the average between the spatial FWHM along x- and y-axis given by the distribution of predicted coordinates. It is interesting to note that the related studies largely utilized smaller and especially thinner monolithic detectors compared to the one used in this thesis (50.8x50.8x20 mm^3). The main disadvantage of using a big monolithic crystal is its cost of production; in addition, the larger the crystal, the more SiPM are needed. This can be solved by combining the readout of rows and columns from a SiPM matrix. M. Stockhoff et al. [37] have proven that the monolith size of 50×50×16 mm^3 coupled with the reduced channels has negligible impact on spatial resolution. Additionally, the greater the volume of the crystal, the longer calibration time is needed. However, the thickness of the detector increases the sensitivity of the PET scanner since the probability of gamma-interaction being detected grows as the thickness of the crystal increases [38].
Besides machine learning algorithms, statistical methods using statistical-based positioning (SBP) and maximum-likelihood [39] [40] have demonstrated high spatial resolution results. P. Fan et al. [33] made a direct comparison of SBP and ANN. The 3D positioning of their ANN model performs better than SBP regarding spatial resolution in x-, y-, and z-directions, where the resolutions are respectively 2.58 mm, 2.56 mm, and 4.79 mm with SBP, and 2.02 mm, 2.03 mm and 3.96 mm with ANN. Additionally, P. Garcia et al. have concluded that ANN is the optimum method when compared with least squares, generalized Chi-Square, and k-NN.

In nearly all TOF-PET detectors, the time-of-interaction is estimated using simple linear methods that measure the time at which the photodetector signal crosses a predefined threshold [41]. However, these methods don't use all the potentially useful timing information contained in the detector waveforms. Machine learning algorithms have been used to overcome this limitation, a common method consists in digitizing the rising edge of the signals and extracting features from it.
One of the earliest investigations of machine learning timing discrimination was described by Leroux et al. [42], who used a multilayer ANN to estimate the temporal information from the digitized photodetector waveforms. Two detectors were compared, an LSO-APD detector and a BGO-APD detector, both tested in coincidence with a fast reference detector used to trigger the waveform digitizer; the neural network timing discriminator provided negligible improvement in timing resolution compared to digital constant fraction discriminator (CFD) [43], possibly because of the low sampling rate used in the study (100 MHz). A successful technique was proven by Berg et al. [44]. They used two LYSO–PMT detectors in coincidence in order to extract two digitized waveforms, which are stored as a 2D vector, where the first dimension is the number of detector channels (two in this case) and the second dimension represents the length of the digitized waveform. This 2D vector was then

fed into a Convolutional Neural Network (CNN) which is a class of ANN, most commonly applied to analyze visual imagery (or arrays). Ground-truth labeled training data were acquired by stepping a gamma source ±7.5 cm about the midpoint between the detectors in 5-mm increments. The results demonstrate that CNN-based TOF estimation improves timing resolution by 20% compared to leading edge discrimination [43] (231 ps versus 185 ps), and 23% compared to CFD (242 ps versus 185 ps).
When dealing with monolithic scintillators coupled with SiPM matrix, temporal information is distributed among all photodetectors, based on the relative position between the scintillation event and the SiPM. Van Dam et al. [45] have developed a maximum likelihood interaction time estimation (MLITE) method to determine the time of interaction from the collection of timestamps obtained by four dSiPM coupled with monolithic LSO:Ce,0.2%Ca detectors of different dimensions. They compared their results with several deterministic methods based on the (weighted) averaging of multiple timestamps. They obtained a CTR of 184 ps for a 24x24x20 mm^3 crystal, an improvement of ∼20%, compared to the method that uses the earliest timestamp only. They concluded that the influence of the OTTS (see Section 1.2.2) of the scintillation photons on the timing performance of monolithic scintillator detectors can at least partially be corrected for by utilizing the information contained in the spatio-temporal distribution of the collection of timestamps.

2. Introduction to Artificial Neural Networks

An artificial neural network (ANN) is a machine learning algorithm that mimics the functioning of biological neurons. The network structure consists of interconnected nodes that are the analogs of brain neurons.
A nerve cell is composed of a cell body and two types of branches: the axon and the dendrites. A neuron receives signal inputs from dendrites, it generates a signal from its cell body and delivers it along the axon. The connection between two neurons is called synapse. When an electric signal passes through a synapse, neurotransmitters are released in order to enhance or inhibit the signal, depending on the nature of the synapse. The synapse effectiveness can be modified by the signal itself; this process allows for the learning process and it is probably responsible for human memory.
The functioning of an ANN depends on its structure and the weighted connections between nodes, like biological neurons, positive weights correspond to excitatory synapses while negative weights model inhibitory ones [46].

2.1. ANN structure

The basic functional unit of ANN is the perceptron, firstly developed in the 1950s and 1960s by the scientist Frank Rosenblatt, inspired by earlier work by Warren McCulloch and Walter Pitts. The function of a perceptron neuron is to take several inputs $\mathbf{x} = x_1, x_2, x_3, \ldots$, and to produce a single binary output. To compute the output the neuron uses a set of weights, real numbers that can be positive, negative, or null. The neuron's output, 0 or 1, is determined by whether the weighted sum $\sum_j w_j x_j$ is less than or greater than some threshold value.

$$output = \begin{cases} 0 & \text{if } \mathbf{w} * \mathbf{x} + b \leq 0 \\ 1 & \text{if } \mathbf{w} * \mathbf{x} + b > 0 \end{cases} \tag{5}$$

An evolution of the perceptron is the sigmoid neuron. The sigmoid neuron is able to produce an output ranging from 0 to 1. This is achieved by computing the output as $\sigma(\mathbf{wx} + b)$, where σ is the sigmoid function, $\mathbf{w}$ and $\mathbf{x}$ are vectors, and $b \equiv -threshold$ is the *bias*.

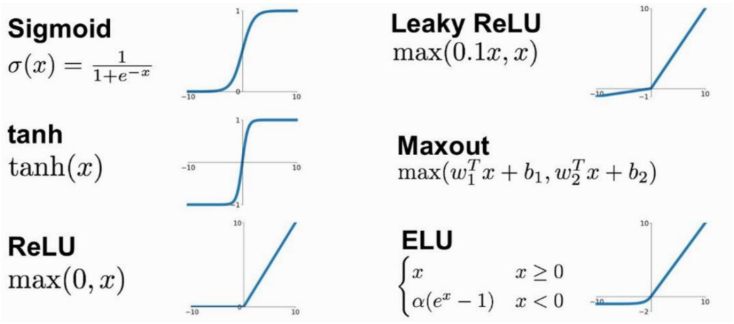


Figure 5: *Some of the most used activation functions [47].*

Generally, the output of a neuron is $u = f(P)$ where $P = \sum_j w_j x_j$ and f is the *activation function*, which is the step function for the perceptron we introduced earlier, and it is the sigmoid function in a sigmoid neuron (examples are shown in Figure 5). Other usually used activation functions are the linear function, the logistic function, and the hyperbolic function [48].
To summarize, each neuron is characterized by the weights associated with each input, the bias (regarded as the weight to be given to a negative unit input), and an activation function used to calculate the output value. The complete structure of an ANN is generally composed of an input layer, one or more hidden layers, and an output layer, each layer is made of one or more nodes.
ANN can be classified based on the connection pattern between their nodes: *feedforward* neural networks (FFNN) have no loop and information goes from the input node to the output node, without going back to previous layers (Figure 6); in this structure, the response to an input is not dependent on the previous state of the network. The most popular class of feed-forward networks is the *multilayer perceptron*, which will be the structure used in this thesis.
Differently, the network is called *recurrent* if it has got at least one feedback loop, for this reason the system is dynamic; these types of architectures are less used because learning algorithms for recurrent nets are (at least to date) less powerful, but they are more similar to how human brains work.

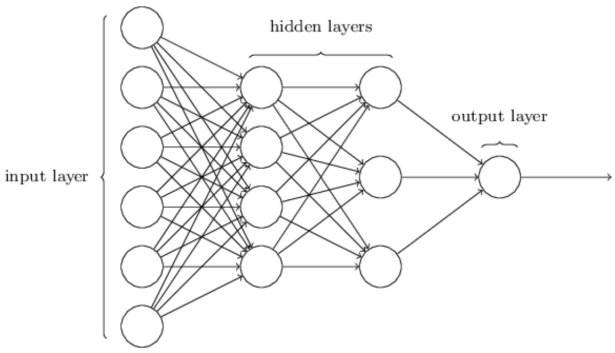


Figure 6: *Schematic of a Feed Forward Neural Network [48].*

2.2. ANN learning

The ability to learn is the fundamental trait of a Neural Network. Interest in ANN derives from their ability to automatically learn from given examples, without a set of rules previously given by human experts. In the ANN contest, the learning process translates into updating the weight of the connections between nodes, and performances are improved by repeating the process iteratively. Being a machine learning method, networks can learn using supervised, unsupervised, or semisupervised learning. There are several algorithms available for each of these methods, including error-correction rules, Boltzmann learning, Hebbian rule, and competitive learning rules. This chapter will focalize on the back-propagation algorithm, which is the basis of the networks that this thesis will discuss next.

Back-propagation is part of error-correction rules, and it is one of the most used learning methods for training feedforward networks. They are called error-correction rules because the network parameters (weights and thresholds) undergo update by iteratively minimizing the difference between the predicted network output and a reference (desired network output), which is provided by the supervisor.
Back-propagation, namely backward propagation of errors, is based on the propagation of the error signals from the output layer back to hidden layers by using the chain rule to iteratively compute gradients for each layer. Training through back-propagation is again an iterative supervised learning process requiring a training set of patterns along with the corresponding desired output set.
The algorithm is divided into two different mechanisms working sequentially at each iteration step. The first mechanism, called direct computation, involves the computation of the neuron output, layer by layer, until a result is generated by the output layer; at the beginning of the learning process, the weights are randomly set to small numbers. The actual output of the network is compared to the expected output for that particular input to obtain an error signal. The second mechanism involves the update of the weights layer by layer working backward from the output layer, through the hidden layer, and to the input layer. This process is repeated until the correct output is produced. Fine-tuning the weights in this way has the effect of teaching the network to produce the correct output for a particular input.
The main issue with updating the weights of the neurons in the hidden layers is that the supervisor does not provide the desired output signal for such neurons, hence an immediate correction based on the desired output

cannot be computed. The solution in the back-propagation is to solve the differential of the error function with respect to the weights of hidden neurons in the function of the weights updated in the next layer by chaining the differential. Therefore, it is implemented by calculating the derivative (gradient) of a *Cost function* (or *Loss function*) with respect to the weights, and then changing each weight by a small increment in the negative (opposite) direction to the gradient. The cost function has the purpose of estimating the error between the network output and the desired one. Choosing the proper cost function depends on the type of the problem to solve; *cross-entropy* function is used to estimate the difference between estimated and predicted probability distributions, hence is more suitable for classification problems, *Mean Squared Error* (MSE) is used for regression problems, where a quantity has to be predicted.

Let's assume we have a multi-layer FFNN with L hidden layers, each layer having M_l neurons and an output layer with N neurons, R is the number of instances in the training dataset. Let $w_{ij}^{(l)}$ be the generic weight of the i_{th} neuron in the layer l with respect to the output of the neuron j in the layer $l-1$.
The factor $\Delta \mathrm{w}_{\mathrm{ij}}^{(\mathrm{l})}$ can be written than as:

$$\Delta \mathrm{w}_{\mathrm{ij}}^{(\mathrm{l})} = \eta \sum_{\mathrm{k}=1}^{\mathrm{R}} \delta_{\mathrm{i}}^{(\mathrm{l}),(\mathrm{k})} y_{\mathrm{j}}^{(\mathrm{l}-1),(\mathrm{k})} \tag{6}$$

where η is the learning rate, and $y_j^{(l-1),(k)}$ is the output signal of the j_{th} neuron of the $(l-1)_{th}$ layer in correspondence of the k_{th} instance of the training set, with:

$$\delta_{\mathrm{i}}^{(\mathrm{l}),(\mathrm{k})} = \begin{cases} \left(t_i^{(k)} - u_i^{(k)}\right) f'\left(P_i^{(k)}\right) & \text{if } l = L \\ \sum_{\mathrm{k}=1}^{\mathrm{M}_{+1}} \left(\delta_{\mathrm{i}}^{(\mathrm{l}+1),(\mathrm{k})} w_{\mathrm{ri}}^{(\mathrm{l}+1)}\right) f'\left(P_{\mathrm{i}}^{(\mathrm{k})}\right) & \text{if } l < L \end{cases} \tag{7}$$

where signals u_i and t_i are the output and the reference value of the i_{th} neuron, M_{l+1} is the number of neurons in the $(l+1)_{th}$ layer, and r is the index of the $(l+1)_{th}$ layer.
The term $f'(P_i^{(k)})$ in equation (7) is the derivative of the activation function computed for $P_i^{(k)} = \sum_j w_{ij} x_j^{(k)}$.
The learning rate η in equation (6) is the parameter that must be tuned in order to balance the quality and speed of convergence.
Equation (6) is derived by using MSE as cost function, which expression is $\sum_{i=1}^{N} \frac{1}{2} (t_i - u_i)^2$.
The overall procedure of the back-propagation algorithm is:

1. Compute the network output with random initial weight for the k_{th} instance
2. Start from $l = L$ (output layer) and compute $\delta_i^{(L),(k)}$ for each neuron i
3. Move to layer $l = L-1$ and compute $\delta_i^{(L-1),(k)}$ in function of $\delta_i^{(L),(k)}$
4. Continue computing $\delta_i^{(l),(k)}$ until $l = 2$
5. Repeat from step (1) to step (4) for all R instances
6. Use equation (6) for all network's weights

During the fitting process, two parameters must be defined: the *batch size* is a hyperparameter that defines the number of samples to work through before updating the internal network weights; the number of epochs is a hyperparameter that defines the number of times that the learning algorithm will work through the entire training dataset, during each epoch, the learning algorithm works on the network updated in the previous epoch. These two numbers must be tuned in order to obtain a solid algorithm [49].

2.3. ANN on Python

There are several platforms for implementing machine learning algorithms. The Artificial Neural Networks discussed in this paper are entirely implemented in Python language by using the TensorFlow platform.
TensorFlow is an open-source software library for machine learning and artificial intelligence. Keras is the high-level API of TensorFlow which provides essential abstractions and building blocks for developing neural networks.
Different python libraries (e.g. NumPy) can be used to apply data preparation which is the first step involved in building a machine learning algorithm. With Keras library, the structure of the ANN can be defined by choosing the number of hidden layers and the number of neurons for each layer (input and output layers included). Subsequently, Keras can be used to train the network defining some training parameters such as the number of epochs and batch size.
Training a network usually involves several trials to find the best combination of both structure and hyperparameters to obtain the best result based on the problem to solve [50].

3. Objective of the research

The main objective of this research is the idealization of two ANNs, one for PET positioning, the other for PET timing. The geometry of the detector under consideration consists of a monolithic crystal of dimensions 50.8x50.8x20 mm^3 coupled to an 8x8 matrix of SiPMs.
Figure 7 shows a schematic representation of the workflow. All data are obtained from simulations performed on the ANTS2 program. From the simulations, the response values of the SiPMs in terms of scintillation photons counts (proportional to the charge produced) are extrapolated, along with the temporal information of each SiPM. The integrated count values are used to train a dedicated positional PET ANN, which provides the two-dimensional position of the scintillation event. These coordinates are later used together with the temporal information to train a second ANN for TOF-PET timing, which outputs a correction value to reduce the temporal uncertainty.

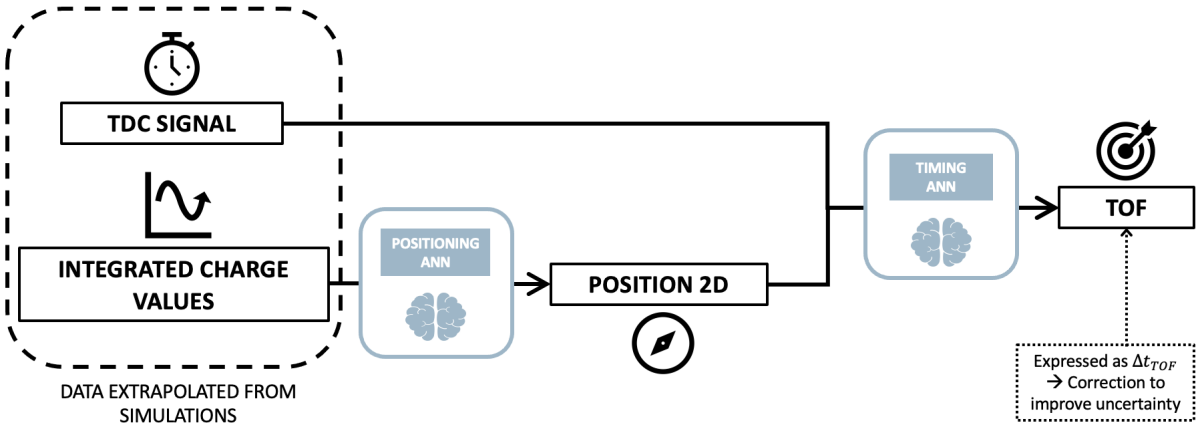


Figure 7: Schematic rapresentation of data workflow.

The details of how the data are processed are set out in the following chapters. Section 4 deals with the first positioning ANN, this network is basically an alternative to linear methods such as center-of-gravity calculation [3], since its purpose is to fit the function that maps the scintillation light distribution to the position-of-interaction. As seen in Section 1.5 several attempts similar to this one have been done, but very few researches were performed on such a thick crystal. Note that only position relative to crystal surface is retrieved, no information on the depth of interaction (DOI) is extrapolated, mainly because of the difficulty of experimentally obtaining a training set for this value. Section 5 deals with the timing ANN that elaborates the temporal information distributed among all SiPMs together with spatial information (coordinates from positioning ANN) in order to improve time resolution. The main idea supporting this method is that the spatial distribution of temporal information given by the SiPM matrix constitutes a feature that an algorithm such as ANN can exploit to correct the temporal uncertainty given by the physics of the detection process; in other words, TOF values are not simply obtained from the first SiPMs that respond, but the event temporal dynamic is exploited to provide better TOF measurements.
This research is set up in such a way that it is then repeatable experimentally, moreover different methods to obtain the training datasets and two different crystals were tested.

4. Positioning ANN

4.1. Materials and methods

4.1.1 ANTS2

ANTS2 is a simulation and data processing package developed for position sensitive detectors with Anger camera type readout. The simulation module of ANTS2 is based on ROOT package from CERN, which is used to store the detector geometry and to perform 3D navigation. The module is capable of simulating particle sources, performing particle tracking, generating photons of primary and secondary scintillation, tracing optical photons and generating photosensor signals [51]. The package includes a time-resolved mode that provides data referred to single time bins.
ANTS2 allows to perform suitable simulations for this research meanly through these tools:

- define type, configuration, and figures of merit of photodetectors
- define scintillation crystal geometry and properties (physical and optical properties)
- define reflective coating materials and their properties
- define a source of 511 keV photons with fixed beam properties
- define beavior of light when passing from one material to another
- define simulation duration and number of time bins with which it is divided (in time-resolved mode)

4.1.2 Setup and simulation parameters

As already mentioned, the photodetectors used in this research are the silicon photomultipliers, specifically the NUV-HD FBK SiPM [52]. Its main figure of merits are reported in Table 3.

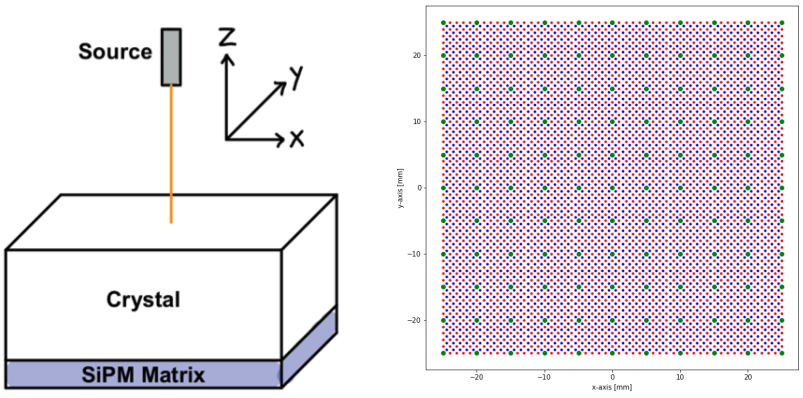
Size $[mm^2]$	PDE	Gain	Recovery time [ns]	Dark rate [Hz]	#SPAD
6.2x6.2	0.45	1	10	$4{\cdot}10^6$	206x206

Table 3: *NUV-HD FBK characteristics.*

The SiPMs are placed in a 8x8 matrix disposition, with a distance of 6.4 mm center to center, hence forming a 51x51 mm^2 array. A 1 mm layer of Optical Grease BC-630 from Saint-Gobain [53] is placed above the SiPM matrix; the optical grease is a transparent, colorless, silicone-based optical coupling compound used to facilitate coupling between SiPM and scintillation crystals. Two crystals and three different configuration are evaluated:

1. **$LaBr_3$(Ce) + Teflon**
   A Cerium-doped Lanthanum(III) Bromide crystal is coated with a 1 mm layer of Teflon. $LaBr_3$(Ce) is incorporated in a 1 mm layer of glass since this crystal is hygroscopic (meaning that exposition to air humidity affects the crystal). Scintillation crystals are usually externally wrapped with reflective materials such as Teflon, in order to convey all scintillation photons towards the photodetectors.
2. **LYSO(Ce) + Teflon**
   A Cerium-doped Lutetium Yttrium Orthosilicate crystal is coated with a 1 mm layer of Teflon. The glass coating is not present since LYSO is not a hygroscopic material.
3. **LYSO(Ce) + ESR**
   Same crystal as point 2 but the coating material is changed to an enhanced specular reflector (ESR) [54], which is an ultra-high reflectivity, mirror-like optical enhancement film.

The main difference between Teflon and ESR is in the light reflection mode: Teflon absorbs 2% of light and scatters back 98% of light following a 180° Lambertian distribution; ESR absorbs 2% of light and the remaining 98% is specularly reflected back.
The gamma source is modeled by a cylindrical volume of 5 mm height and 1 mm diameter, output photons have an energy of 511 KeV and they are collimated in order to form a pencil beam perpendicularly entering the crystal surface.



(a) *Detector geometry recreated in the simulation environment, center of the reference system is located at the center of the crystal.*

(b) *Training points locations of grid 51x51 (red), grid 50x50 (blue), and grid 11x11 (green).*

Figure 8: *Setup and training points distribution.*

The data needed to build training and test datasets were collected by translating the source along the x-y plane (see Figure 8a for the reference system), following a grid of points, at each point 500 events were recorded. Three datasets were collected corresponding to three different grid configurations, as shown in Figure 8b:

- grid 51x51: composed by 51x51 points, 1 mm spacing between points, grid vertices are at coordinates (±25,±25)
- grid 50x50: composed by 50x50 points, 1 mm spacing between points, grid vertices are at coordinates (±24.5,±24.5)
- grid 11x11: composed by 11x11 points, 5 mm spacing between points, grid vertices are at coordinates (±25,±25)

Three ANNs were trained using different dataset combinations: grid 51x51 was initially used alone to test a training points spacing of 1 mm, secondly, the 51x51 grid was merged with the 50x50 grid to obtain a resulting training points spacing of $\sqrt{2}/2$ mm, finally the 11x11 grid was used to evaluate the effect of a reduced number of training points (suitable for a faster experimental calibration).

4.1.3 Data preparation

Data exported from ANTS2 consists in the photon counts for each SiPM (computed for the entire event duration when time-resolved option is off), together with coordinates and delivered energy of each interaction in the crystal. Matlab [55] was used to group data from all points into a table, subsequently, the `.csv` file created was imported into Spyder [56] (a Python 3.7.11 compiler) on which the rest of the analysis continued.
The main steps in preparing the dataset for the positioning ANN are as following:

- Data are filtered based on the total counts detected by the SiPM matrix, this step is required in order to discard events during which the gamma photon did not fully interact with the crystal. Filtering is performed by fitting the total counts' histogram with a gaussian function, the mean and FWHM of the curve are computed, data outside the interval $[mean - FWHM, mean + FWHM]$ are discarded (see Figure 9a.
- A realistic interaction position is computed: ANTS2 provides first, secondary and following interaction positions of a single incoming gamma photon, together with the energy delivered for each of them, hence the signal detected by the SiPMs based on their location can be better represented by the energy-weighted position of interaction p_w, which is computed with the following formula:

$$p_w = \frac{\sum_i \mathbf{x_i} \cdot E_i}{\sum_i E_i}$$

  where $\mathbf{x_i}$ is the vector of coordinates of the i_{th} interaction and E_i is its delivered energy. This coordinate is mostly used to achieve more accurate data placement during data analysis.
- Each event provides 64 signals coming from the 8x8 SiPM matrix, this number can be reduced to 16 values by summing signals along each row and each column; combining the readout of rows and columns preserves the positional information and it reduces the number of input values to the ANN.
- Data are normalized over the highest value present in the dataset: this is a common procedure for preparing a training dataset for ANNs.
- 30% of total data is used as test dataset, the 80% of remaining is used as training dataset, the remaining is used as validation dataset.

4.2. Results

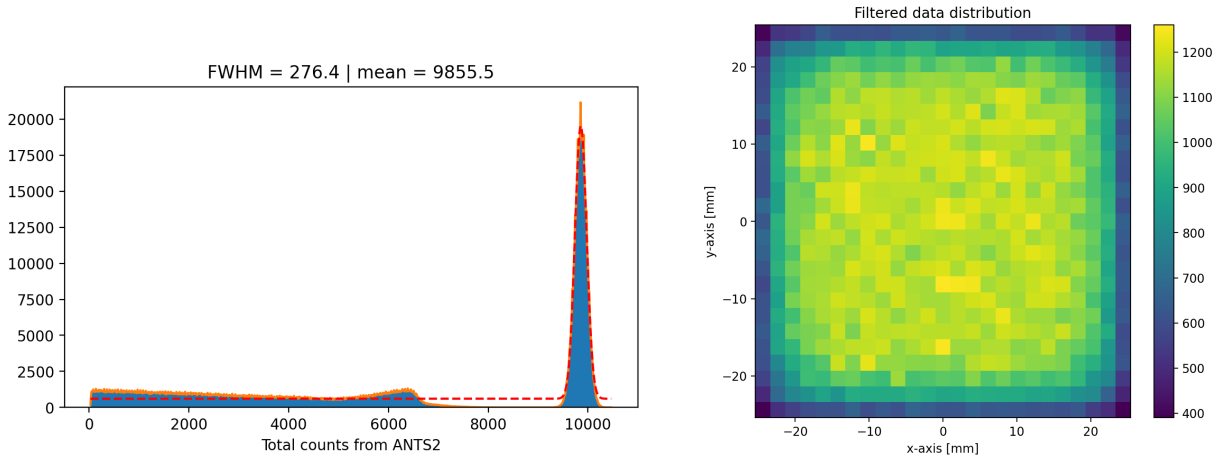
4.2.1 Data analysis

Before implementing the network, some properties of the collected data were analyzed, in order to understand the characteristics of the scintillation photons and how SiPMs respond to them. Note that the results presented in this section refer to the $LaBr_3$(Ce) + Teflon configuration irradiated by the 51x51 grid.
Figure 9a shows the distribution of the total counts detected from the SiPM matrix, this variable coincides with the sum of the counts of each SiPM per event. The total counts are directly related to the total energy detected through Formula (8) in fact, it is the parameter with which experimentally the data are usually filtered.

$$E[KeV] = \frac{N_{phe}}{Y \cdot PDE} \tag{8}$$

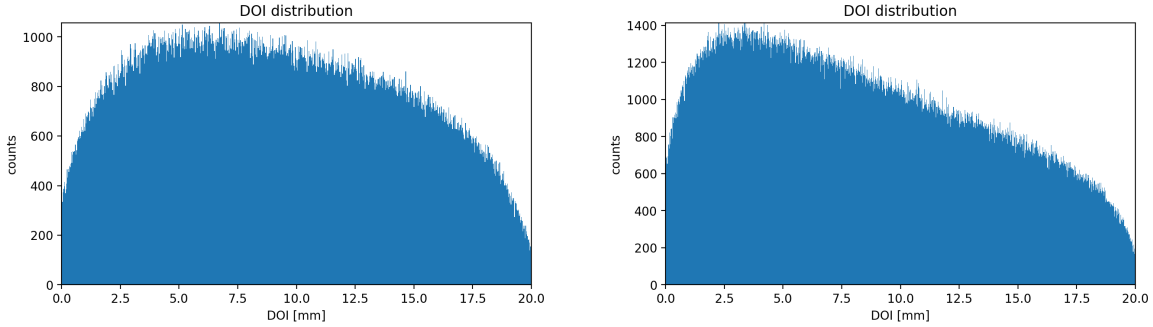
As already mentioned, the events that form the Gaussian peak are those that almost totally delivered the 511 KeV of the gamma photon. These events are retained after filtering and they account for 48.7% of the original data. Filtered data distribution based on their p_w are shown in Figure 9b; as expected, events interacting near

the crystal's edge are more likely to escape the crystal and to not deliver their full amount of energy, hence they are less present in the filtered dataset.



(a) *Total counts histogram fitted with a Gaussian function.* (b) *Filtered data distribution.*

Figure 9: *Data analysis for $LaBr_3(Ce)$ + Teflon configuration.*



(a) *DOI distribution in $LaBr_3(Ce)$.* (b) *DOI distribution in LYSO(Ce).*

Figure 10: *DOI distribution comparison between the two crystals.*

Depth of interaction (DOI) distribution of the filtered data is reported in Figure 10, the distribution is coherent with the densities and the correlated attenuation coefficients for both crystals: in Figure 10b, the distribution has more events with low DOI than the one in the Figure 10a, this is because of the higher attenuation coefficient of LYSO(Ce) than that of $LaBr_3$(Ce) (0.86 cm^{-1} vs. 0.47 cm^{-1}).

4.2.2 Positioning ANN results

Two main figures of merit are used to present results on positioning ANN:

- *Spatial error* is the euclidian distance between the given label (coordinates of source position) and the predicted coordinates. It expresses the magnitude of the error made by the network.
- *Spatial FWHM* is given by a Gaussian fitting applied on the spatial distribution of the predicted coordinates, around the entry point. It is computed for both x- and y- axes, and the average value of the two is reported. FWHM measures the spatial resolution of the detected image.

Results are usually reported as the mean spatial error and the mean spatial FWHM, they are computed among all positions over the crystal. All results are obtained by feeding the trained positioning-ANN with data from the test dataset.

Each scintillator configuration was trained with both grid 51x51 and grid 11x11. Only $LaBr_3$(Ce) + Teflon was also trained with the combination of 51x51 grid + 50x50 grid in order to understand how an increased resolution of the training dataset affects the ANN's performance.

Grid	mean spatial error [mm]	mean spatial FWHM [mm]
51x51	1.807	2.107
51x51 + 50x50	1.704	2.074

Table 4: *Training grid comparison using $LaBr_3(Ce)$ + Teflon.*

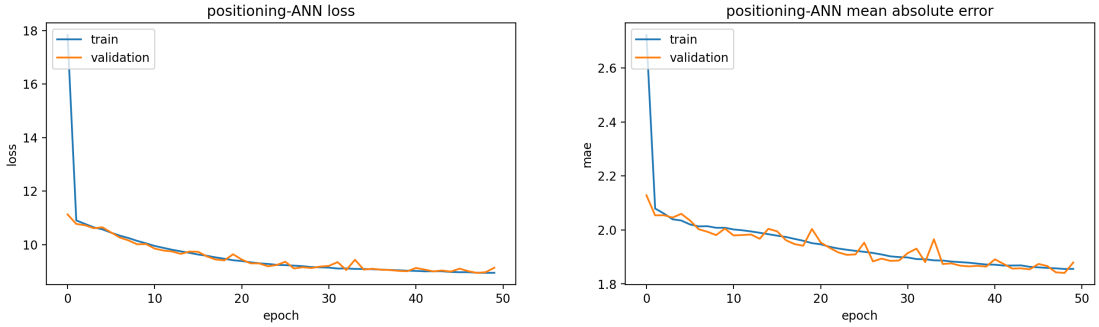
Several ANN networks were tested, the best performances were obtained by structures with hidden layers composed of a number of nodes equal to powers of 2. In Table 5, performances of the best two structures are shown for each crystal. The hidden layer structure is investigated since input and output layers are the same for each positioning ANN. Given these results, subsequent analyses are done using the 256-256 structure for $LaBr_3$(Ce) and the 256-256-128-64 structure for LYSO(Ce), which seem to perform slightly better for the two crystals.
All investigated networks use the ReLu function (see Figure 5) as activation function for the nodes in the hidden layers; a linear activation function is always used in the output layer since these ANNs perform a regression. The chosen loss function is the mean squared error.

	$LaBr_3$(Ce) + Teflon		LYSO(Ce) + Teflon	
ANN structure[1]	mean spatial error [mm]	mean spatial FWHM [mm]	mean spatial error [mm]	mean spatial FWHM [mm]
256-256	1.807	2.107	1.583	2.237
256-256-128-64	1.806	2.198	1.538	2.195

[1] Each value refers to the number of nodes for that hidden layer, input and output layers are not included.

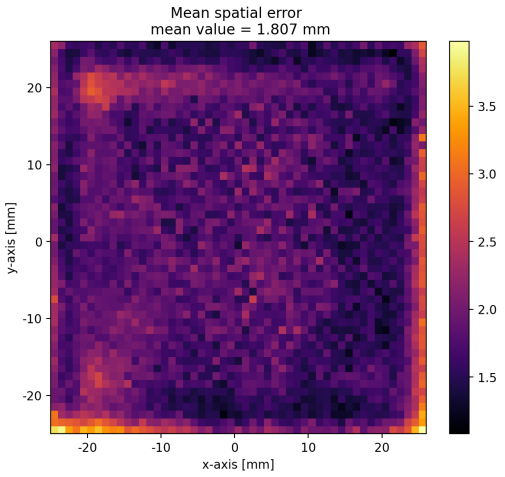
Table 5: *Comparison on different ANN's structures. 51x51 grid was used.*

Figure 11 shows the trend of loss function and mean absolute error, for both training and validation sets. Values are plotted in function of the epoch (see Section 2.2).

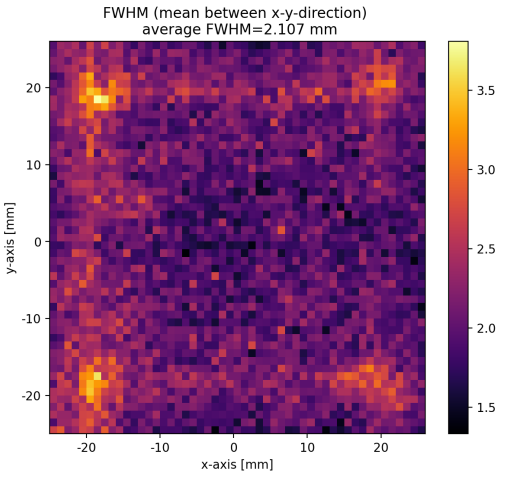


(a) *Learning curve.* (b) *Error curve.*

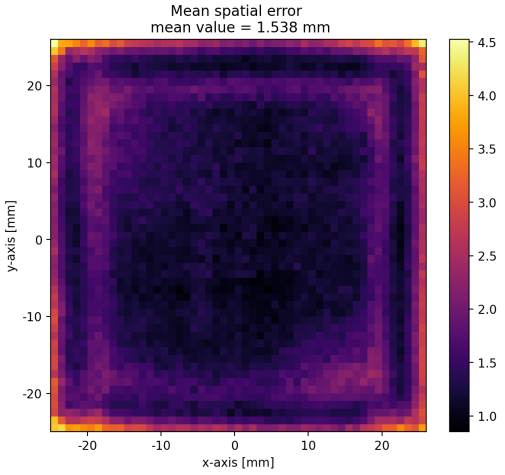
Figure 11: *Fitting results for $LaBr_3(Ce)$ + Teflon / 51x51 grid*



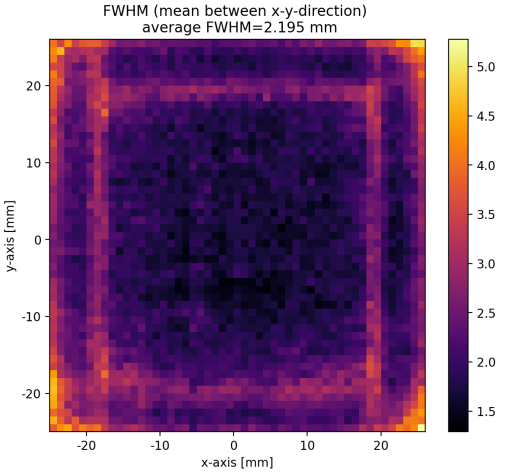
(a) *Spatial error / $LaBr_3(Ce)$ + Teflon.*



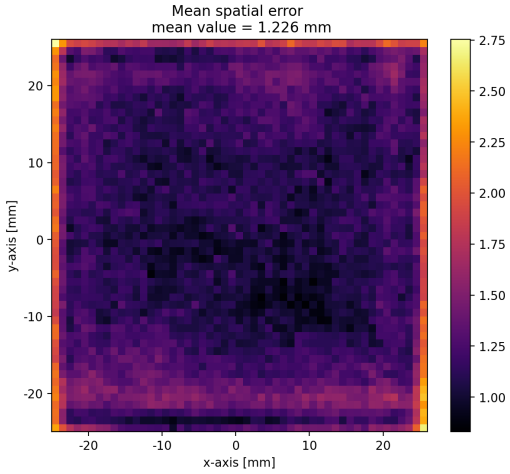
(b) *Spatial FWHM / $LaBr_3(Ce)$ + Teflon.*



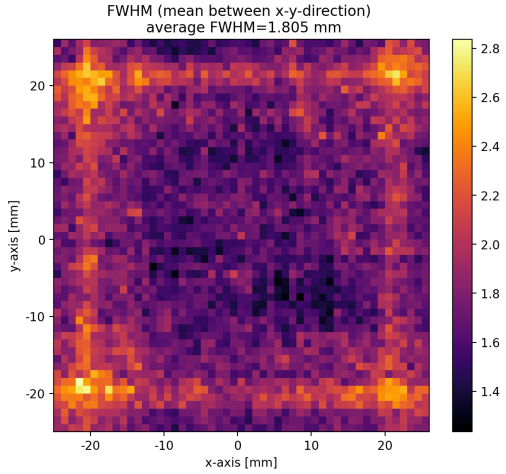
(c) *Spatial error / LYSO(Ce) + Teflon.*



(d) *Spatial FWHM / LYSO(Ce) + Teflon.*

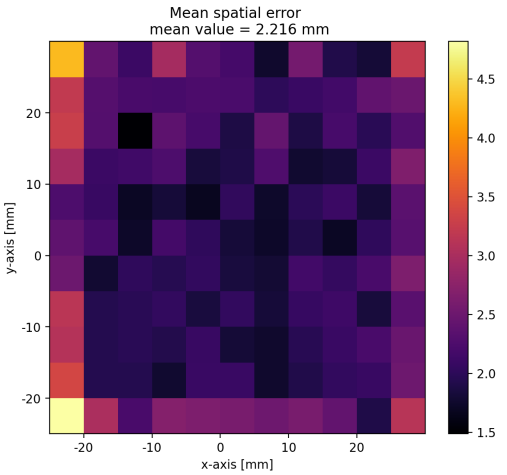


(e) *Spatial error / LYSO(Ce) + ESR.*

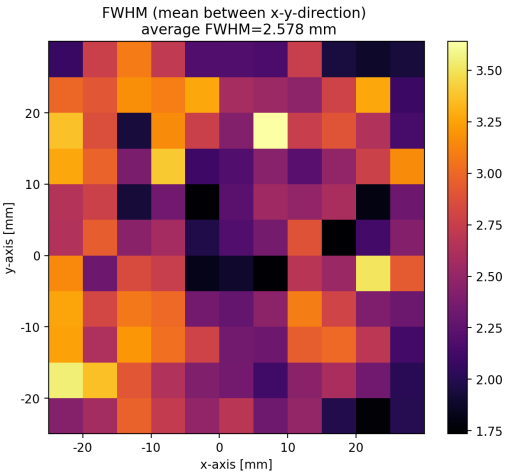


(f) *Spatial FWHM / LYSO(Ce) + ESR.*

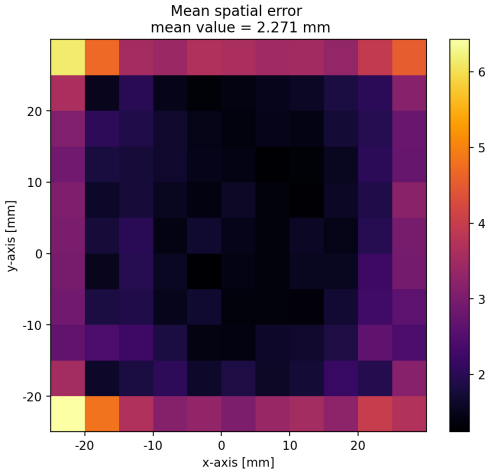
Figure 12: *Results for training with 51x51 grid*



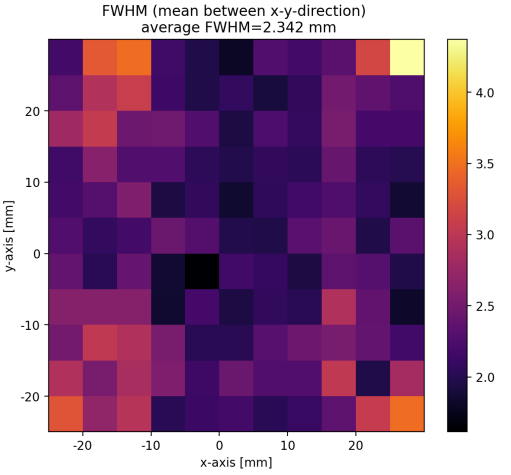
(a) *Spatial error / $LaBr_3(Ce)$ + Teflon.*



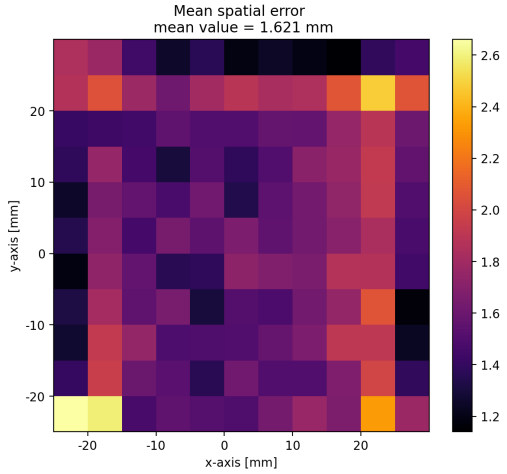
(b) *Spatial FWHM / $LaBr_3(Ce)$ + Teflon.*



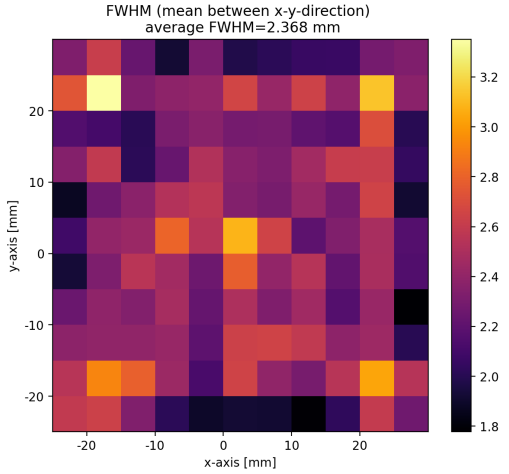
(c) *Spatial error / LYSO(Ce) + Teflon.*



(d) *Spatial FWHM / LYSO(Ce) + Teflon.*



(e) *Spatial error / LYSO(Ce) + ESR.*



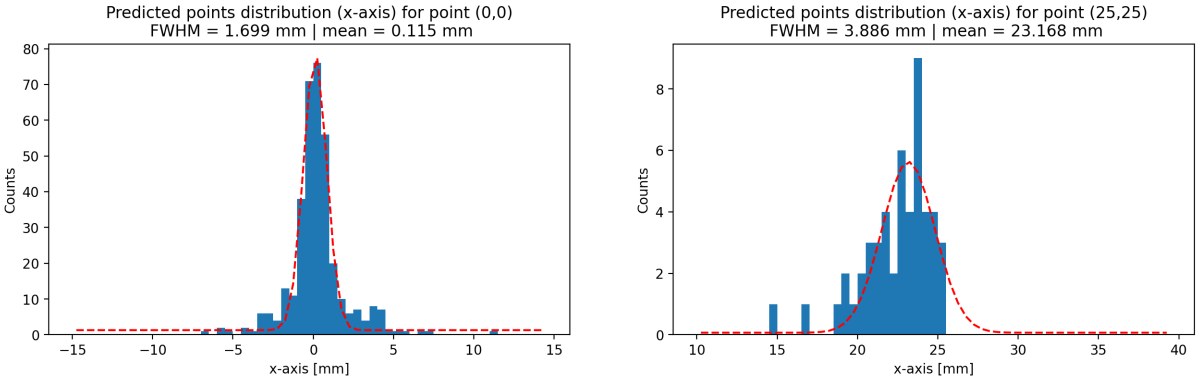
(f) *Spatial FWHM / LYSO(Ce) + ESR.*

Figure 13: *Results for training with 11x11 grid*

Configuration	51x51 grid		11x11 grid	
	mean spatial error [mm]	mean spatial FWHM [mm]	mean spatial error [mm]	mean spatial FWHM [mm]
$LaBr_3(Ce)$ + Teflon	1.807	2.107	2.216	2.578
LYSO(Ce) + Teflon	1.538	2.195	2.271	2.342
LYSO(Ce) + ESR	1.226	1.805	1.612	2.368

Table 6: *Positioning ANNs results comparison.*

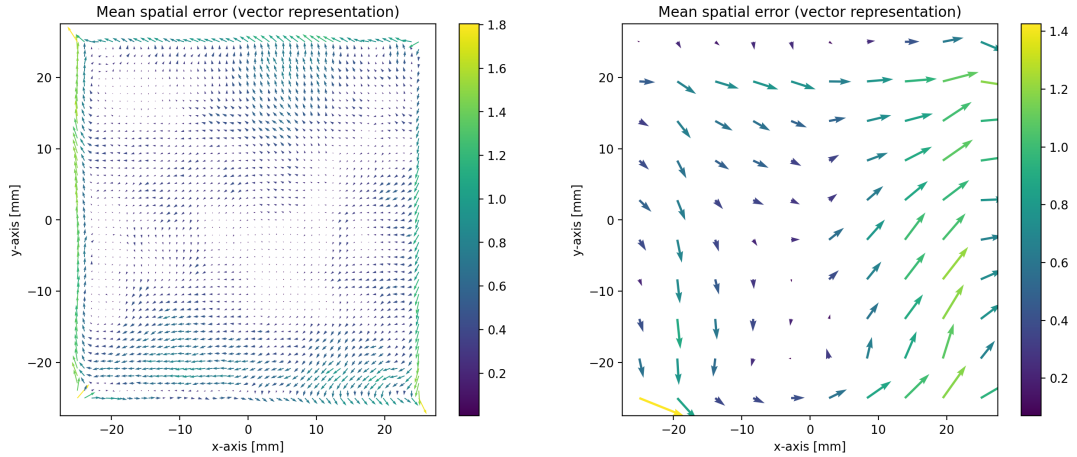
Figures 12 and Figures 13 are images representing the magnitude of the corresponding figure of merit based on the position along the x-y plane; the mean spatial resolution is computed among all points inside a square of evaluation, each test point has its square of evaluation (hence Figures 12 are composed by 1x1 mm^2 squares, and Figures 13 by 5x5 mm^2 squares); spatial FWHM values refers to the spatial distribution around each point of the test grid.
Table 6 shows the average values of the two figures of merit over all the area, for the 6 different combinations of configurations under consideration.



(a) *Point (0,0) (center of the crystal).* (b) *Point (25,25) (vertex of the crystal).*

Figure 14: *Spatial distribution of predicted points along the x-axis. LYSO(Ce) + Teflon | 51x51 grid*

Figure 14 shows the spatial distribution along the x-axis of the predicted points, in two different locations on the crystal. These graphs are explanatory of how the spatial FWHM was calculated.

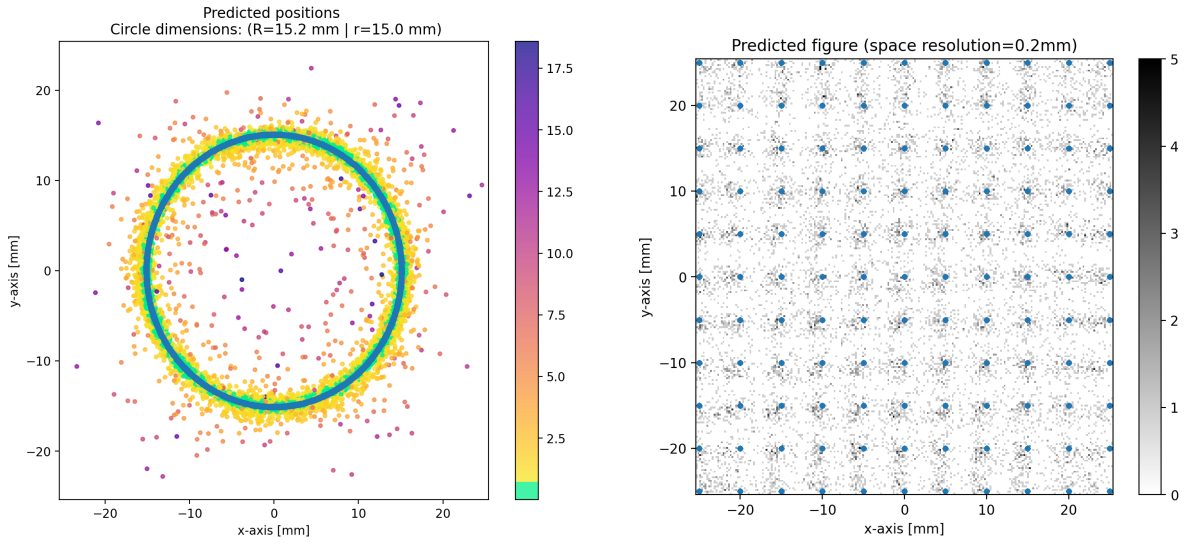


(a) *LYSO(Ce) + ESR | 51x51 grid.* (b) *LYSO(Ce) + ESR | 11x11 grid.*

Figure 15: *Vector representation of the mean spatial error.*

The spatial error represents a distance in a 2D space (in fact, it is the difference between two vectors), hence it

has a magnitude and a direction; a bidimensional representation of the mean spatial error is shown in Figure 15, each arrow refers to the mean spatial error calculated around the training grid point used, the colour is related to the magnitude.



(a) *Prediction of a circle (blue points). Colours represent the spatial error in [mm], green points have a spatial error under the mean value.*
LYSO(Ce) + ESR | 51x51 grid

(b) *Prediction of the 11x11 grid (blue points). Greyscale represent the count of predicted points in a 0.2x0.2 mm^2 square.*
LYSO(Ce) + ESR | 11x11 grid

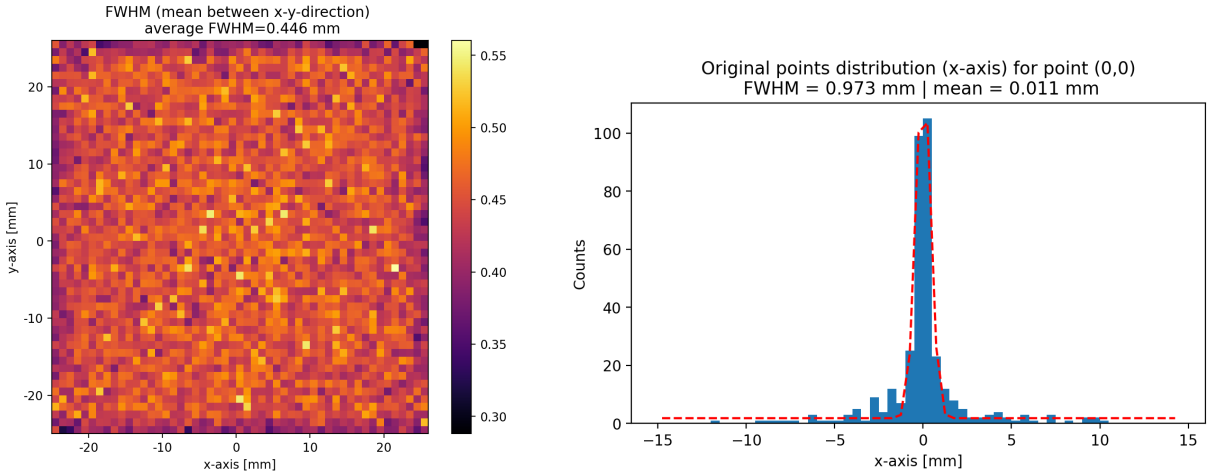
Figure 16: *Study on predicted images.*

Figure 16 is an attempt to show how the positioning ANNs predict a given image. To form Figure 16a, data from grid 51x51 were used, subsequently, all events with p_w inside a fixed region were kept (blue points), the region is such that to form a circle with an internal diameter of 15 mm and an external one of 15.2 mm, finally, the predictions for these points were plotted. Figure 16b shows a different type of visualization, it consists of a 2D histogram of predicted coordinates from the 11x11 grid, darker spots contain more counts.

4.3. Discussion

The fitting results shown in Figure 11 demonstrate a good fitting of the ANN over the training dataset. Both training and validation plots in Figure 11a decrease to a point of stability with a minimal gap between the two final loss values, this gap is relatively small, proving that the training data is representative of the problem. The curves flatten out toward the end of the graphs, suggesting that more epochs would not significantly improve performance.
Table 4 shows that there is a relatively small improvement when the spatial resolution of the training dataset is increased. The improvement is not so high to justify the large increase in calibration time that one would have to use to collect 51x51+50x50 points instead of 51x51 points. Concerning the ANN structure, the number of hidden layers seems to not affect the results very much, however, there is to consider that in early stages of this research, training using energy-weighted position p_w as label (see Section 4.1.3) showed that structure 256-256 performs better than structure 256-256-128-64 for $LaBr_3$(Ce) + Teflon. This might suggest the need to test the most appropriate architecture when changing the combination of the parameters involved in the problem. The ANN's number of nodes is directly reflected in the complexity of the network, an aspect to be evaluated if the network must be implemented in hardware such as FPGA [57].
From Figure 12 and Figure 13 the spatial dependency of ANN performances is evident. The spatial resolution is best at the center of the crystal and starts to degrade toward the edges. It is interesting to note (especially in Figure 12 that has a higher resolution) that both spatial error and spatial FWHM get worse in an intermediate region, at about 5-7 mm far from each edge. This effect is due to the reflection of scintillation light at the sides of the crystal, which affects the light distributions on the SiPM matrix making it increasingly similar for nearby interaction positions toward the detector edges (as if the matrix detected two sources). Thus, the ability of the ANN to predict the correct position generally deteriorated by nearly 50% among all configurations. The same edge-effect has already been experienced by Seifert et al. [58], who suggest that the influence of light distribution similarity might be minimized by intelligent photosensor design with smaller pixels toward the edges of the crystal.

The performances degradation at the edges of the crystal can be easily understood from Figure 14b which, in comparison with Figure 14a, shows a worst Gaussian distribution of data, since the ANN has difficulties in evaluating a light distribution that misses points because of the end of the crystal.
The results shown in Table 6 are in line with previous research on such thick crystals. For example, Borghi et al. [59] obtained a 1.7 mm spatial FWHM with a 32x32x22 mm^3 LYSO(Ce) crystal, using slightly bigger SiPMs (8x8 mm^2).
As expected performances are better when the training spatial resolution is higher, however it should be considered that a considerable reduction in training points corresponds to a relatively small worsening. The best configuration is the LYSO(Ce) + ESR; LYSO(Ce) performs generally better than $LaBr_3$(Ce), and the mean spatial error is the one that benefits most from the transition from Teflon to ESR.
There is to consider that the results on spatial FWHM are affected by two factors: the source pencil beam diameter of 1 mm, and the Compton scattering, since the gamma photon can be deviated and can deliver its energy in secondary interactions, resulting in a p_w coordinate different from the entering x-y position; this effect is shown in Figure 17b, where the distribution of the energy-weighted positions p_w along the x-axis is spread around the entry point. This effect is rather uniformly present throughout the detector (see Figure 17a, since it comes from the Compton scattering. It is possible to compare Figure 17b with Figure 14a to infer that the network actually adds 0.762 mm to the spatial FWHM in point (0,0).



(a) *Spatial FWHM of p_w distribution over the crystal surface.*

(b) *p_w distribution along x-axis for point (0,0)*

Figure 17: *p_w analysis for LYSO(Ce) + Teflon*

Finally, the consequences of the spatial FWHM can be better appreciated in Figure 16, in particular, Figure 16b is the visualization of the detector 2D point spread function (PSF), directly related to how the detector detects point figures, in this case, the 11x11 grid.

5. Timing ANN

This section focuses on the training of a timing ANN whose goal is to improve time resolution. As discussed in Section 1.2.2, the temporal distribution $p_{t_d}(t \mid \Theta)$ is responsible for a certain distribution of detector detection times, when multiple events are recorded in the same condition, this results in a Gaussian distribution of TOF values with a certain temporal FWHM.
The main idea supporting this method is that the spatial distribution of temporal information given by the SiPM matrix constitutes a feature that an algorithm such as ANN can exploit to correct the temporal uncertainty given by the physics of the detection process; in other words, TOF values are not simply obtained from the first SiPMs that respond, but the event temporal dynamic is exploited to provide better TOF measurements. This is possible due to the fact that we are using a monolithic scintillator in which light can propagate freely, it would not be possible in a segmented scintillator.
As already mentioned, this research is intended to be in support of a laboratory experiment, hence a temporal coincidence experiment was modeled to train the network, and data from ANTS2 were processed precisely to simulate this setup. Then another setup was simulated to test the success of the entire workflow that is the goal of this research (see Section 3).

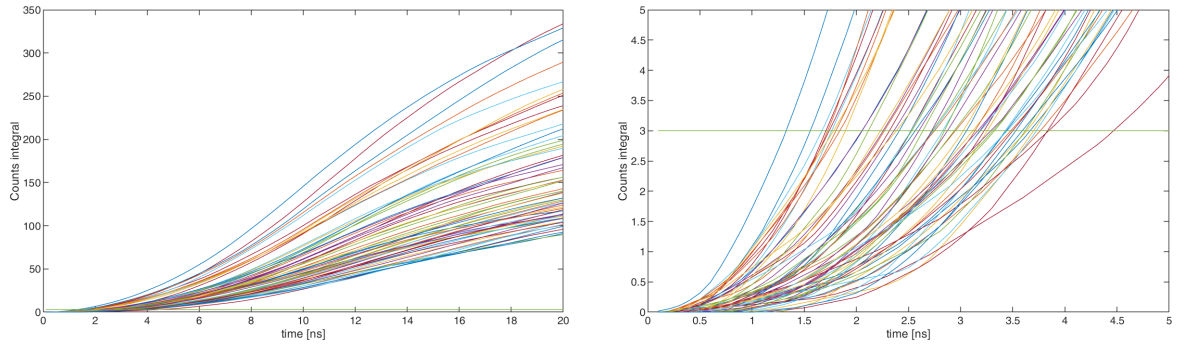
5.1. Materials and methods

5.1.1 Setup and simulation parameters

Data acquisition for the timing ANN is the same as the one described in Section 4.1.2 for the positioning ANN. However, for this temporal analysis, only grid 11x11 was used, and only $LaBr_3$(Ce) + Teflon and LYSO(Ce) + Teflon configurations were tested.
ANTS2 simulations were performed by turning on the time-resolved option; since the analysis focuses on the first temporal instants of detection, a 20 ns window was chosen as the length of the simulation. ANTS2 data starts from the scintillation instant of the event, no information from source emission to scintillation instant is recorded. The 20 ns length of the simulation was divided into 200 time bins, obtaining a time resolution of 0.1 ns. This time resolution is a compromise between simulation accuracy and the size of the data produced (some exceed 10 Gigabytes).

5.1.2 Data preparation

Data exported from ANTS2 consists of the photon counts for each SiPM, interaction coordinates, and relatively delivered energy, but in this case, each line refers to a single time bin, hence each event is composed of a number of lines equal to the number of time bins.
Data were elaborate in Matlab, where the time profile of the response of each SiPM was obtained. Every event was assigned 64 timestamps, coming from the crossing of a threshold by the integrated time response of each SiPM; the response of a SiPM is the number of counts of scintillation photons as a function of time. Its integration over time gives the graphs in Figure 18.



(a) *Graph for the entire simulation.*

(b) *Zoom from the previous figure. An example of a threshold at 3 photons is also represented.*

Figure 18: *Integrated photon counts of 64 SiPMs for a single event. Each count corresponds to a single photon detected by a SiPM.*

Data must first be used by the positioning ANN (as explained in Section 3), hence they undergo the same data processing listed in Section 4.1.3.
Some events can present missing values for one or more timestamps, this occurs because, after a scintillation, not all SiPMs have a sufficient response to pass the threshold to get the timestamp; this phenomenon will be referred to as *under-threshold*. The problem is that the ANN's input layer expects always the same number of input values for each instance that has to evaluate, hence two possible solutions were evaluated:

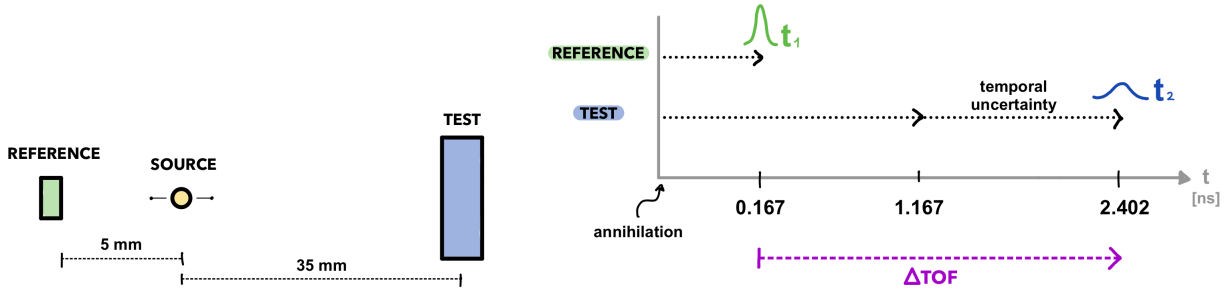
- each under-threshold is assigned the value of the last possible timestamp (20 ns, since it is the duration of the simulation)
- each under-threshold is assigned the value of the maximum timestamp recorded during that event

The second option was chosen, since the temporal distribution is better preserved in this way, rather than assigning a value dependent on an arbitrary variable such as the simulation duration.

	% of events with at least one under-threshold	
	Threshold = 1 phe	Threshold = 3 phe
$LaBr_3(Ce)$ + Teflon	7.73%	11.43%
LYSO(Ce) + Teflon	14.86%	17.70%

Table 7: *Percentages of under-threshold events for different crystals and thresholds.*

5.1.2.1 From simulation to experiment



(a) *Experimental setup.* (b) *Temporal dynamic scheme.*

Figure 19: *Reproduction of an experimental setup.*

In order to simulate a real experiment, data from ANTS2 need to be processed. The idea is to reproduce the experimental setup shown in Figure 19a. A reference detector is placed coincident with the test detector, the gamma source is placed in between, at distances of 5 mm and 35 mm respectively.
The first step is to create data from an ideal reference detector using a Gaussian distribution with chosen mean and FWHM. The reference detector used from Borghi et al. [25] was used as a model. It is a 3x3x5 mm^3 LSO(Ce) detector, which presents a time resolution of ∼89 ps. This value is used as FWHM of the modeling Gaussian distribution; obviously, the curve's mean coincides with the travel time of the gamma photon from source to reference detector (0.167 ns).
Subsequentially, data from simulations, which refers to the ones provided by the test detector, must be shifted temporally to include the time delay due to the gamma photon traveling from the source to the test detector (1.167 ns), along with the time of flight within the crystal before interacting with it (related to DOI). A schematic representation of the temporal dynamic of events is shown in Figure 19b.
The data obtained so far are the time measurement t_2 of the test detector and that of the reference detector t_1. Both values represent the time between annihilation and the crossing of the threshold by the signal of the first responding SiPM. This quantity will generally be referred to as TOF.
Both t_1 and t_2 refer to the exact annihilation instant, which is something that can not be retrieved experimentally. The time data that would be obtained from this experimental setup would be the value of the test detector measurement referred to the ones of the reference detector. For this reason the difference between t_2 and t_1 is computed to obtain ΔTOF. A visual representation of this variable is shown in Figure 19b. Note that each coincidence event is associated with a ΔTOF, since each event from the test detector was randomly matched with an event from the reference detector, in order to simulate a coincidence event.

5.1.2.2 Timing ANN label

As already mentioned, if the detector was perfect and not subject to temporal uncertainty given by $p_{t_d}(t \mid \Theta)$, the TOF distribution would be centered in a single value. In reality, the temporal uncertainty increases the width of the distribution. In order to improve TOF measurements, we want the timing ANN to learn the correction needed to tighten the distribution and to reduce uncertainty, hence the difference between the recorded ΔTOF and the mean value between all ΔTOFs. This quantity will be called *correction* Δt and provides the label to be assigned to the i_{th} event.
Given what was discussed above, it was chosen to define the correction Δt as:

$$\Delta t_i = \Delta TOF_i - \frac{\sum_j^N \Delta TOF_j}{N} \tag{9}$$

where N is the total number of events used for training the timing ANN.
The correction $\Delta t_{predicted}$ predicted by the network (after training it) can be then used to correct the relative ΔTOF in order to obtain a $\Delta TOF_{corrected}$.

5.1.2.3 Timing ANN input

When dealing with experimental data, each SiPM provides a timestamp value that is affected by gamma-photon time travel, distance between detectors, and measurement global clock. To get a performing method regardless of where the data is centered, we need to train the network with data that is not dependent on these factors but easily retrievable to be given to the network during TOF-PET imaging, regardless of all the conditions that may vary between measurements.
If all timestamps are subtracted the minimum of timestamps for that event, the relative delay between the first and other timestamps is obtained. These values preserve the temporal dynamic that the ANN must use to retrieve the correction, meanwhile, they are not referred to any offset. A visualization of how *normalized timestamps* are obtained is reported in Figure 20.

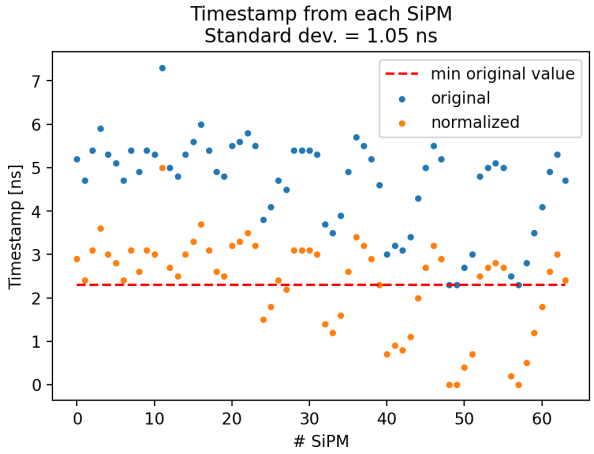


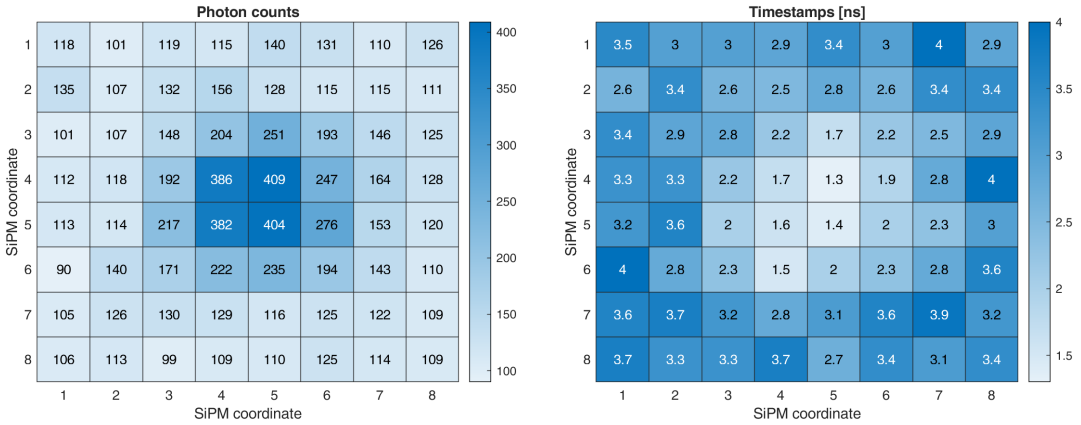
Figure 20: *Original and normalized timestamps of each SiPM for a single event.*

As previously mentioned, timestamps are coupled to the 2D coordinates of the interaction to form the data to be given as input to the network, which will then receive 64+2 values.

5.2. Results

5.2.1 Data analysis

The following temporal data analysis was performed using the $LaBr_3$(Ce) + Teflon configuration, and it was irradiated with the 11x11 grid, but this time 1000 events were recorded in each source position. A threshold of 3 photons was chosen to collect timestamps for the data analysis.

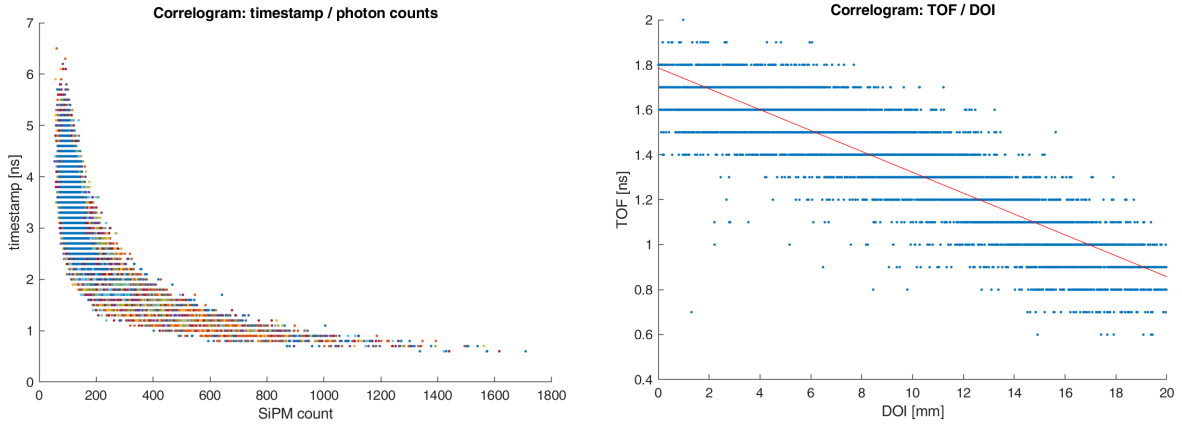


(a) *Photon counts recorded from the SiPM matrix for a single event.*

(b) *Timestamps recorded from the SiPM matrix for a single event.*

Figure 21: *Photon counts and timestamps distribution over the SiPM matrix. Source position is placed over the center of the detector.*

The total amount of photons detected by each SiPM with the relative timestamp are shown in Figure 21. As expected, they are clearly related: SiPMs closer to the scintillation point have a higher photon count and a lower timestamp, meaning that they respond more and faster.

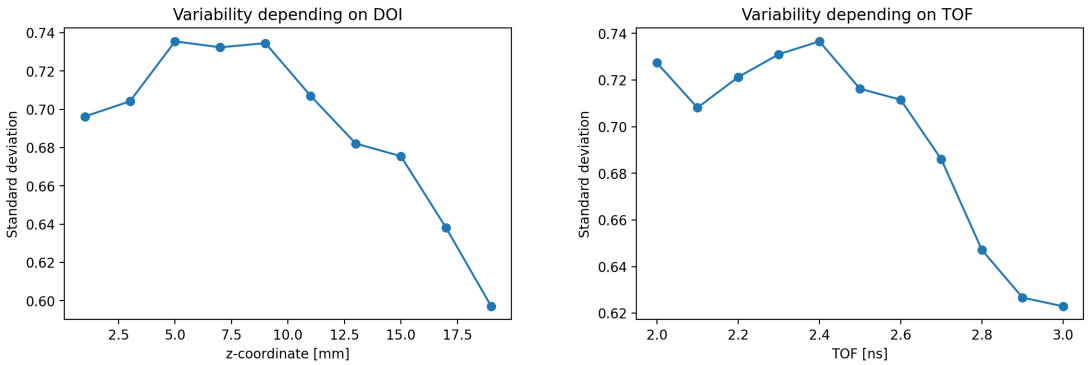


(a) *Correlogram between photon counts and timestamps. Each point represent the response of one SiPM referred to one event.*

(b) *Correlogram between TOFs and DOIs values. Each point represents a single event.*

Figure 22: *Correlograms.*

Figure 22a is a confirmation of the inverse proportionality that generally exists between the count and the timestamp of each SiPM. From Figure 22b, the first timestamp collected (i.e. the TOF) is linearly dependent on the DOI of the event, since scintillations occurring deep in the crystal take less time to be detected by the SiPM array, in contrast, those far from the photodetectors have to travel a longer distance.

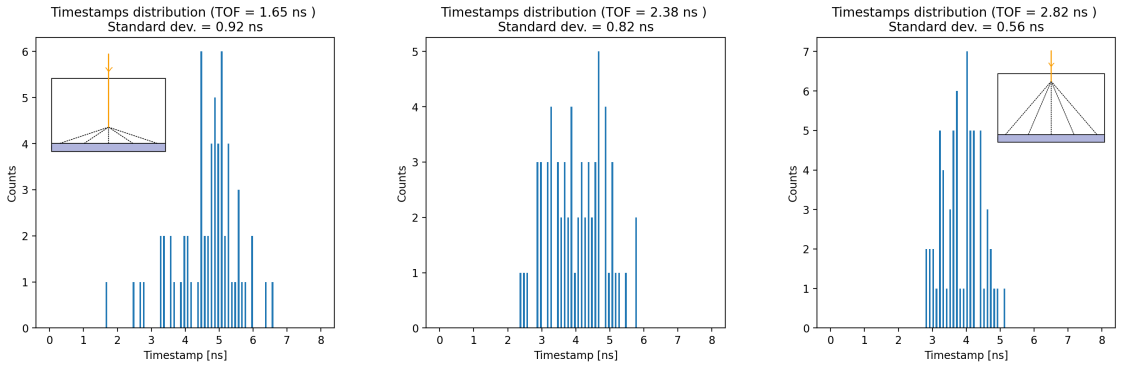


(a) *DOI-dependent variability. Z=0 represent the interface between the SiPMs and the crystal.*

(b) *TOF-dependent variability.*

Figure 23: *Variability dependencies.*

Figure 23 presents a study on the variability of the 64 timestamps given by a single event, in function of its DOI and TOF. Standard deviation was used to quantify the variability of each sample. In Figure 23a, the crystal was divided into 10 sections each 2 mm thick, for each section the average of the standard deviation of the 64 timestamps was calculated. For Figure 23b, on the other hand, the partitioning was done according to the TOF of the event, dividing the distribution into 11 classes.
Taking into account Figure 22b, it is not surprising that the trend of variability as a function of DOI is similar to that as a function of TOF. It is precisely Figure 23a that suggests that the temporal distribution of timestamps is linked to the related TOF, and thus may help ANN to make the right correction.



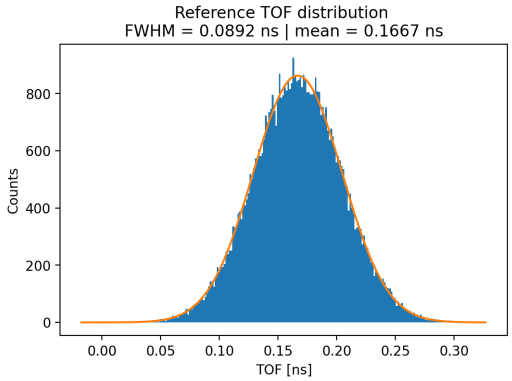
(a) *Event with high timestamps variability and low TOF.*

(b) *Event with moderate timestamps variability and medium TOF.*

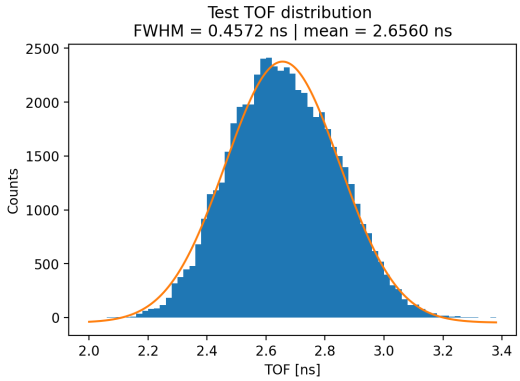
(c) *Event with low timestamps variability and high TOF.*

Figure 24: *Examples of timestamps distributions with different TOFs.*

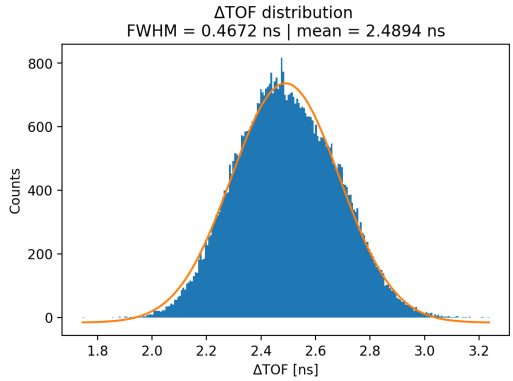
Figure 24 is intended again to demonstrate the concept expressed in the previous paragraph: the distribution of timestamps is related to the relative TOF, comparing the three figures shows that as TOF increases, the distribution of timestamps tightens. The figure is correlated by two drawings showing where the variability in timestamps comes from: when scintillation occurs close to the photodetectors, photons arrive at relatively very different times on the SiPMs (Figure 24a), in contrast, arrival times are very similar for scintillations with high TOF and low DOI (Figure 24c).
The Δt corrections needed for the three events in Figure 24 are respectively -0.557 ns, 0.164 ns, and 0.603 ns. It is from this that one can infer that the sign and absolute value of the correction are related to the temporal distribution of timestamps. This relationship is the one that the timing ANN should model. The addition of the two-dimensional interaction coordinate should help the network recognize particular effects due to proximity to detector edges.



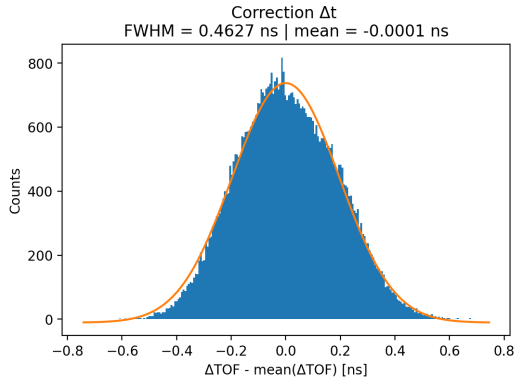
(a) *Distribution of TOFs from the reference detector.*



(b) *Distribution of TOFs from the test detector.*



(c) ΔTOF *distribution.*



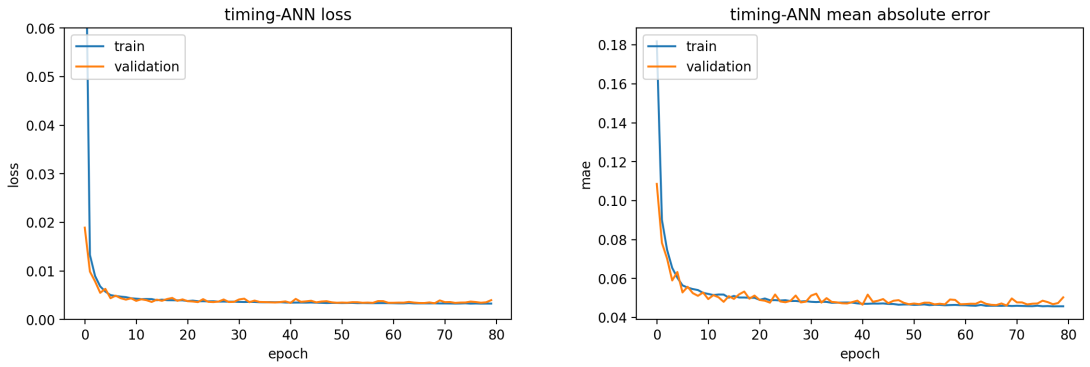
(d) Δt *distribution.*

Figure 25: *Distribution of data from the simulated coincidence experiment.*

Figure 25 includes the distribution from all data obtained by the simulated coincidence experiment (procedure explained in Section 5.1.2.1). The delay due to the distance between the source and the point of scintillation has already been added to the data in Figure 25b. The inherent uncertainty of the test detector is found to be 457.2 ps, this increases for ΔTOF values since they come from the coincidence of the two detectors, as if the reference detector added a part of uncertainty. Obviously, the Δt distribution coincides with the ΔTOF distribution, but the first one is centered in 0 ns.

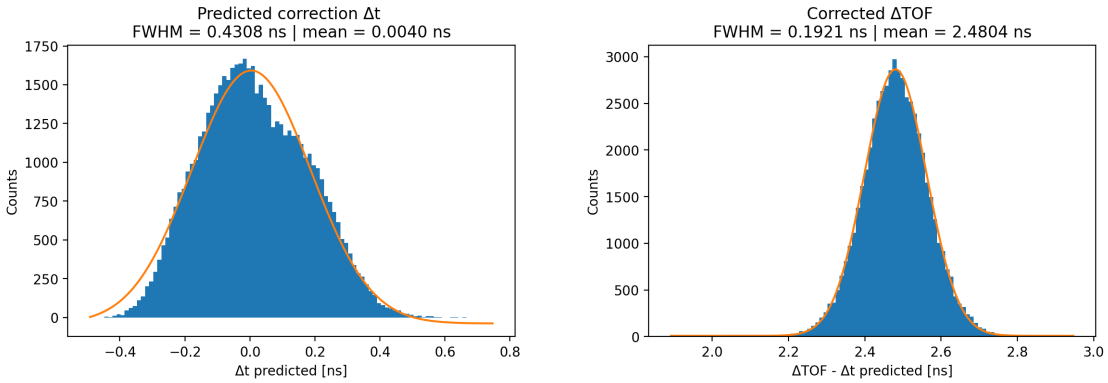
5.2.2 Timing ANN results

The following results on ANN timing were obtained with the configuration $LaBr_3(Ce)$ + Teflon; the 11x11 grid was used as the training dataset and a threshold of 3 photons was set up to collect the timestamps. The positioning ANN presented in Section 4 was used to predict the input coordinates to the timing ANN. The timing ANN structure that best corrects the ΔTOF distribution consists of two hidden layers, made of 64 and 32 nodes respectively. Also, this network uses the ReLu function as activation function; the output layer presents a linear activation function since also this problem is a regression. The chosen loss function remains the mean squared error.



(a) *Learning curve.* (b) *Error curve.*

Figure 26: *Fitting results for $LaBr_3(Ce)$ + Teflon | 11x11 grid.*



(a) *Predicted corrections Δt distribution.* (b) *Corrected $\Delta TOFs$ distribution.*

Figure 27: *Timing ANN results.*

Figure 27a represents the predicted data coming directly from the timing ANN, i.e. the value of correction Δt that is associate to that event. Figure 27b is nothing more than the distribution of the corrected ΔTOFs, computed using Formula (10).
It is worth pointing out that Figure 27a and Figure 27b are distributions, hence they are composed of the values associated with each event that forms the test dataset. For example, an event having ΔTOF = 2.163 ns (placed in the left tail of the distribution of Figure 25c) is associated with a $\Delta t_{predicted}$ = -0.379 ns (left tail of Figure 27a), the resulting $\Delta TOF_{corrected}$ is 2.542 ns, which is placed at the center of the distribution in Figure 27b. Applying Formula (10) to all events in the test dataset allows us to derive Figure 27.

$$\Delta TOF_{corrected} = \Delta TOF - \Delta t_{predicted} \tag{10}$$

The figure of merit used to quantify the timing ANN's performance is the *Improvement* (I) defined as:

$$I = \frac{FWHM_{original}}{FWHM_{corrected}} \cdot 100\%$$

where $FWHM_{original}$ refers to the original ΔTOFs distribution, and $FWHM_{corrected}$ refers to the corrected ΔTOFs distribution. The root mean square error (rmse) is also computed between original ΔTOF values and their mean value (the same was done for the $\Delta TOF_{corrected}$).

	$LaBr_3$(Ce) + Teflon		LYSO(Ce) + Teflon	
	Threshold = 1 phe	Threshold = 3 phe	Threshold = 1 phe	Threshold = 3 phe
$FWHM_{original}$ [ps]	282	467	340	561
$FWHM_{corrected}$ [ps]	189	192	246	272
Improvement	149%	243%	138%	206%
$rmse_{original}$ [ps]	124	183	144	245
$rmse_{corrected}$ [ps]	84	85	108	118

Table 8: *Timing ANN results for different setups and thresholds.*

Table 8 shows the results for the two configurations, each with thresholds set at 1 and 3 photons.

5.2.3 Validation

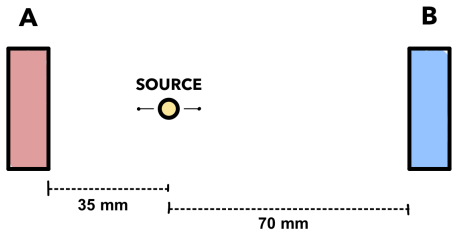


Figure 28: *Coincidence experiment setup.*

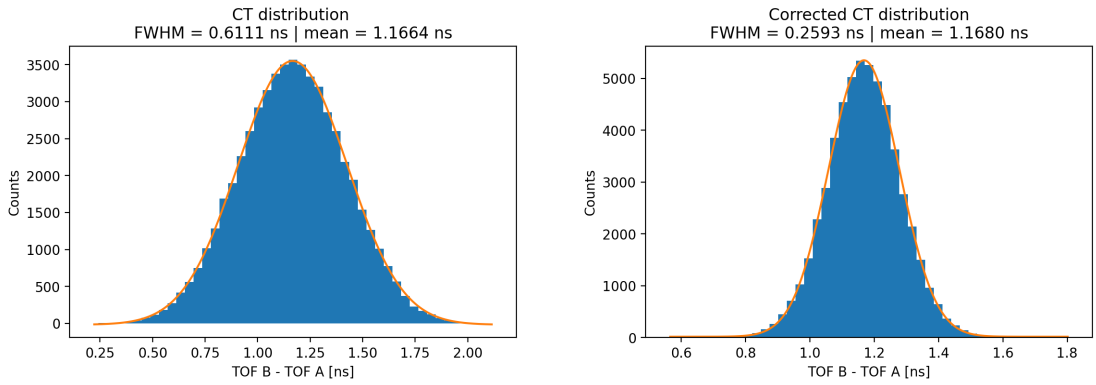
The goal of this last part of the research is to test the improvement of TOF-PET that can be obtained using the combination of the positioning and timing ANNs. To do this, data were collected from two separate simulations (both with $LaBr_3$(Ce) + Teflon and 3 photons threshold); these were used to recreate the coincidence experimental setup shown in Figure 28. The goal is to compare the distribution of the *coincidence time* (CT), which is the value needed during TOF-PET in order to improve imaging (see Section 1.2.1).

$$CT = TOF_B - TOF_A \tag{11}$$

The relative time delay between source and detector was added to each detector, according to their distance of 35 mm and 75 mm. Similar to what was done during training between test and reference detectors, events from detector A were randomly matched with events from detector B, in order to simulate coincidence events.
Initially, the difference between the time signals of the two detectors was calculated using the uncorrected TOFs (equation (11)). Finally, $CT_{corrected}$ was computed using the corrected TOFs from both detectors. Note that the correction procedure is applied independently on events from detector A and detector B, only at the end the *corrected* TOF_B and the *corrected* TOF_A are placed in coincidence (subtracted).
The procedure can be summarized as follows:

1. 64 photon counts + 64 timestamps are collected from each event.
2. Events are filtered based on total counts (energy-filtering).
3. Under-threshold events are assigned the value of the maximum timestamp for that event.
4. For each event, the 64 photon counts are normalized, turned into 16 signals (row-column sum), and fed to the positioning ANN.
5. For each event, normalized timestamps and predicted coordinates are given to the timing ANN.
6. For each event, the predicted correction Δt is used to correct the TOF.

Figure 29 compares the results on CT obtained with and without correction.



(a) *CT distribution.* (b) $CT_{corrected}$ *distribution.*

Figure 29: *Comparison between corrected and no-corrected CT.*

5.3. Discussion

The fitting results shown in Figure 26 demonstrate a good fitting of the timing ANN over the training dataset. The trend of the curves is similar to that obtained for the positioning ANN (see Section 4.3), the data appear to be representative of the problem, and a sufficient degree of prediction is obtained, without the need to increase the number of epochs.
Results presented in Figure 27 are promising. The distribution of the corrected ΔTOF has a FWHM of 192.1 ps, against the one of the original ΔTOF which is 467.2 ps, meaning that temporal data uncertainty is corrected considerably. It should not be forgotten that the data presented are distributions, which means that among the adjusted ΔTOFs there are still events for which the temporal uncertainty deviates them from the mean (those toward the tails of the distribution), but the magnitude of this effect is much smaller than for the unadjusted ΔTOFs; this means that the probability of finding different time measurements from each other (for same detector and the same distance from the source) is reduced when the timing ANN is used.
Given equation (9), it can be demonstrated that the corrected ΔTOF distribution (Figure 27b) coincides with the error made by the network in predicting the Δt, except for a constant:

$$error = \Delta t - \Delta t_{pred.} = \Delta TOF - \frac{\sum_j^N \Delta TOF_j}{N} - \Delta t_{pred.} = \Delta TOF_{corrected} - \frac{\sum_j^N \Delta TOF_j}{N} \tag{12}$$

Since the mean $\frac{\sum_j^N \Delta TOF_j}{N}$ is a constant value, equation (12) shows that the uncertainty of the network in making the prediction is directly reflected in the uncertainty of the final outcome. Hence, by improving the timing ANN's capability of prediction, temporal uncertainty is directly reduced.
Is to note that, the results obtained depend only on the FWHM of the ΔTOF distribution, which is an intrinsic property of the test detector (partly conditioned by that of the reference detector used during training), results do not depend on where the distribution is centered; this means that the proper correction is applied even if the time data have different delays due to sources at different distances, as happens during TOF-PET imaging, since the spatial delay is only a constant. Once the network is trained on a particular detector, it is ready to correct TOFs unless the intrinsic variability worsens, e.g. due to factors such as environmental influences on photodetectors or electronics; it would be interesting to study this dependence through the experimental application of this research.
With regard to the data presented in Table 8, it can be seen that the configuration $LaBr_3$(Ce) + Teflon generally gives better results, this is not surprising given the physical properties of the two crystals (see Table 1). $LaBr_3$(Ce) has a faster response due to its short decay time, together with a higher light yield which improves the SiPMs response. The 1 photon threshold gives better results than the 3 photons one, but the ANN has a greater improvement on the latter. This could be due to the greater variability of the integrated responses of SiPMs as they develop over time (see Figure 18), this is related to the chosen time resolution of the simulation, which does not allow accurate collection of timestamps for a relatively low threshold. It is noteworthy that the integrated response and not the instantaneous response of the SiPM was chosen precisely to emphasize the slope of the response with its intensity, thus its closeness to scintillation point. However, the choice of threshold must be made considering the noise of the readout electronic in use (state of art electronics are able to detect a single photon and even less). When performing a laboratory experiment, false counts due to dark counts are to be taken into account; a possible solution to reduce them consists in cooling down the SiPMs, additionally, they can be discarded using, for example, a double validation threshold, or by filtering in energy.

Finally, the validation part of this research (Section 5.2.3) gives an idea of the potential of the ANNs built. Figure 29 shows a 236% improvement between the corrected CTR and the no-corrected one. Both distributions are centered almost exactly at the correct temporal delay (∼1.675 ns), but the corrected one is less spread out, in fact, more than 95% of the measurements differ by less than 220 ps from the expected value.

6. Conclusions

The current thesis work aimed to provide a new method to exploit ANN's capabilities for the benefit of monolithic PET and TOF-PET imaging systems. The research focuses on two main objectives: using a positioning ANN to predict the scintillation position and a timing ANN to improve the detector time resolution.
Through simulations, it was possible to train a positioning ANN with good spatial resolution on relatively thick crystals (50.8x50.8x20 mm^3). Results for different configurations were compared, such as different training datasets, different crystals, and coatings; the best setup showed a 1.805 mm spatial FWHM.
The second part showed that is possible to correct temporal information by giving the timing ANN all timestamps collected by a SiPM matrix. From the analysis of the temporal data and the results of the ANN, the time delay between timestamps appears to be the information capable of making the proper correction to the detector's inherent temporal uncertainty. For the setup on which the network performs best, we went from a TOF uncertainty of 467 ps FWHM to one of 192 ps FWHM; furthermore, by simulating a coincidence setup, it was shown that the network can theoretically improve the coincidence time resolution from 611 ps to 259 ps.
The next step is to validate this thesis through an experimental test in the laboratory. Future development can aim to improve the ANN's performances, by testing new structures or using different hyperparameters. Regarding the setup, it would be interesting to extend the variability of the training datasets by using events incident to the crystal at an angle other than 90 degrees. As well as exploiting crystals with instantaneous scintillation, such as the ones that emit Cherenkov photons [60]. Finally, increasing the simulation time resolution could lead to better performances of the timing ANN since the timestamp variability could be better appreciated also for lower thresholds.

References

[1] Zainab T Al-Sharify, Talib A Al-Sharify, Noor T Al-Sharify, et al. A critical review on medical imaging techniques (ct and pet scans) in the medical field. In *IOP Conference Series: Materials Science and Engineering*, volume 870, page 012043. IOP Publishing, 2020.

[2] Emil Jan Konopinski. Beta-decay. *Reviews of Modern Physics*, 15(4):209, 1943.

[3] Kuang Gong, Eric Berg, Simon R Cherry, and Jinyi Qi. Machine learning in pet: from photon detection to quantitative image reconstruction. *Proceedings of the IEEE*, 108(1):51–68, 2019.

[4] Daniel W Rickey, Richard Gordon, and Walter Huda. On lifting the inherent limitations of positron emission tomography by using magnetic fields (magpet). *Automedica*, 14(1):355–369, 1992.

[5] `https://www.tudelft.nl/tnw/over-faculteit/afdelingen/radiation-science-technology/research/research-groups/medical-physics-technology/research/tof-pet`. Accessed: 2022-05-23.

[6] Charles L Melcher. Scintillation crystals for pet. *Journal of Nuclear Medicine*, 41(6):1051–1055, 2000.

[7] Todd E Peterson and Lars R Furenlid. Spect detectors: the anger camera and beyond. *Physics in Medicine & Biology*, 56(17):R145, 2011.

[8] Glenn F Knoll. *Radiation detection and measurement*. John Wiley & Sons, 2010.

[9] Dennis R Schaart, Georg Schramm, Johan Nuyts, and Suleman Surti. Time of flight in perspective: instrumental and computational aspects of time resolution in positron emission tomography. *IEEE transactions on radiation and plasma medical sciences*, 5(5):598–618, 2021.

[10] Christian Wolfgang Fabjan. Collider detectors for multi-tev particles. Technical report, 2002.

[11] Nicoleta Dinu. Silicon photomultipliers (sipm). In *Photodetectors*, pages 255–294. Elsevier, 2016.

[12] Claudio Piemonte and Alberto Gola. Overview on the main parameters and technology of modern silicon photomultipliers. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 926:2–15, 2019.

[13] Fabio Acerbi and Stefan Gundacker. Understanding and simulating sipms. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 926:16–35, 2019.

[14] Stefan Gundacker and Arjan Heering. The silicon photomultiplier: fundamentals and applications of a modern solid-state photon detector. *Physics in Medicine & Biology*, 65(17):17TR01, 2020.

[15] GL Brownell, CA Burnham, S Wilensky, S Aronow, H Kazemi, and D Strieder. New developments in positron scintigraphy and the application of cyclotron-produced positron emitters. In *Medical Radioisotope Scintigraphy. VI Proceedings of a Symposium on Medical Radioisotope Scintigraphy*, 1969.

[16] Suleman Surti, Austin Kuhn, Matthew E Werner, Amy E Perkins, Jeffrey Kolthammer, and Joel S Karp. Performance of philips gemini tf pet/ct scanner with special consideration for its time-of-flight imaging capabilities. *Journal of Nuclear Medicine*, 48(3):471–480, 2007.

[17] Maurizio Conti and Bernard Bendriem. The new opportunities for high time resolution clinical tof pet. *Clinical and Translational Imaging*, 7(2):139–147, 2019.

[18] Thomas F Budinger. Time-of-flight positron emission tomography: status relative to conventional pet, 1983.

[19] Dennis R Schaart. Physics and technology of time-of-flight pet detectors. *Physics in Medicine & Biology*, 2021.

[20] Joshua W Cates, Stefan Gundacker, Etiennette Auffray, Paul Lecoq, and Craig S Levin. Improved single photon time resolution for analog sipms with front end readout that reduces influence of electronic noise. *Physics in Medicine & Biology*, 63(18):185022, 2018.

[21] Davide Marano, Giovanni Bonanno, Salvatore Garozzo, Alessandro Grillo, and Giuseppe Romeo. New improved model and accurate analytical response of sipms coupled to read-out electronics. *IEEE Sensors Journal*, 16(1):19–21, 2015.

[22] Ting-Yi Yang. Machine learning for high resolution 3d positioning of. 2019.

[23] Zhi Li, Mateusz Wedrowski, Peter Bruyndonckx, and Gerd Vandersteen. Nonlinear least-squares modeling of 3d interaction position in a monolithic scintillator block. *Physics in Medicine & Biology*, 55(21):6515, 2010.

[24] Stefan Gundacker, Rosana Martinez Turtos, Etiennette Auffray, Marco Paganoni, and Paul Lecoq. High-frequency sipm readout advances measured coincidence time resolution limits in tof-pet. *Physics in Medicine & Biology*, 64(5):055012, 2019.

[25] Giacomo Borghi, Valerio Tabacchini, Stefan Seifert, and Dennis R Schaart. Experimental validation of an efficient fan-beam calibration procedure for k-nearest neighbor position estimation in monolithic scintillator detectors. *IEEE Transactions on Nuclear Science*, 62(1):57–67, 2015.

[26] Jerome H Friedman. Stochastic gradient boosting. *Computational statistics & data analysis*, 38(4):367–378, 2002.

[27] Florian Müller, David Schug, Patrick Hallen, Jan Grahe, and Volkmar Schulz. Gradient tree boosting-based positioning method for monolithic scintillator crystals in positron emission tomography. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 2(5):411–421, 2018.

[28] Herman T Van Dam, Stefan Seifert, Ruud Vinke, Peter Dendooven, Herbert Lohner, Freek J Beekman, and Dennis R Schaart. Improved nearest neighbor methods for gamma photon interaction position determination in monolithic scintillator pet detectors. *IEEE Transactions on Nuclear Science*, 58(5):2139–2147, 2011.

[29] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

[30] Fumio Hashimoto, Kibo Ote, Ryosuke Ota, and Tomoyuki Hasegawa. A feasibility study on 3d interaction position estimation using deep neural network in cherenkov-based detector: A monte carlo simulation study. *Biomedical Physics & Engineering Express*, 5(3):035001, 2019.

[31] P Conde, A Iborra, AJ González, L Hernández, P Bellido, L Moliner, JP Rigla, MJ Rodríguez-Álvarez, F Sánchez, M Seimetz, et al. Determination of the interaction position of gamma photons in monolithic scintillators using neural network fitting. *IEEE Transactions on Nuclear Science*, 63(1):30–36, 2016.

[32] Y Wang, W Zhu, X Cheng, and D Li. 3d position estimation using an artificial neural network for a continuous scintillator pet detector. *Physics in Medicine & Biology*, 58(5):1375, 2013.

[33] P García De Acilu, I Sarasola, M Canadas, R Cuerdo, P Rato Mendes, L Romero, and C Willmott. Study and optimization of positioning algorithms for monolithic pet detectors blocks. *Journal of Instrumentation*, 7(06):C06010, 2012.

[34] Mateusz Wędrowski. *Artificial neural network based position estimation in positron emission tomography*. PhD thesis, Ph. D. dissertation, Vrije Universiteit Brussel, 2010.

[35] Peter Bruyndonckx, Cedric Lemaitre, DJ Van Der Laan, Marnix Maas, Dennis Schaart, Wang Yonggang, Zhi Li, M Krieguer, and Stefaan Tavernier. Evaluation of machine learning algorithms for localization of photons in undivided scintillator blocks for pet detectors. *IEEE Transactions on Nuclear Science*, 55(3):918–924, 2008.

[36] Peter Bruyndonckx, Sophie Léonard, Stefaan Tavernier, Cedric Lemaître, Olivier Devroede, Yibao Wu, and Magalie Krieguer. Neural network-based position estimators for pet detectors using monolithic lso blocks. *IEEE Transactions on Nuclear Science*, 51(5):2520–2525, 2004.

[37] Mariele Stockhoff, Roel Van Holen, and Stefaan Vandenberghe. Optical simulation study on the spatial resolution of a thick monolithic pet detector. *Physics in Medicine & Biology*, 64(19):195003, 2019.

[38] Peter Bruyndonckx, Cedric Lemaître, Dennis Schaart, Marnix Maas, Magalie Krieguer, Olivier Devroede, Stefaan Tavernier, et al. Towards a continuous crystal apd-based pet detector design. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 571(1-2):182–186, 2007.

[39] Xiaoli Li, Cate Lockhart, Tom K Lewellen, and Robert S Miyaoka. A high resolution, monolithic crystal, pet/mri detector with doi positioning capability. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2287–2290. IEEE, 2008.

[40] T Ling, TK Lewellen, and RS Miyaoka. Depth of interaction decoding of a continuous crystal detector module. *Physics in Medicine & Biology*, 52(8):2213, 2007.

[41] JM Monzo, R Esteve, Christoph Werner Lerche, N Ferrando, J Toledo, RJ Aliaga, V Herrero, and FJ Mora. Digital signal processing techniques to improve time resolution in positron emission tomography. *IEEE Transactions on Nuclear Science*, 58(4):1613–1620, 2011.

[42] Jean-Daniel Leroux, Marc-AndrÉ Tetrault, Daniel Rouleau, Catherine M Pepin, J-B Michaud, Jules Cadorette, RÉjean Fontaine, and Roger Lecomte. Time discrimination techniques using artificial neural networks for positron emission tomography. *IEEE Transactions on Nuclear Science*, 56(3):588–595, 2009.

[43] Suleman Surti and Joel S Karp. Advances in time-of-flight pet. *Physica medica*, 32(1):12–22, 2016.

[44] Eric Berg and Simon R Cherry. Using convolutional neural networks to estimate time-of-flight from pet detector waveforms. *Physics in Medicine & Biology*, 63(2):02LT01, 2018.

[45] Herman T Van Dam, Giacomo Borghi, Stefan Seifert, and Dennis R Schaart. Sub-200 ps crt in monolithic scintillator pet detectors using digital sipm arrays and maximum likelihood interaction time estimation. *Physics in Medicine & Biology*, 58(10):3243, 2013.

[46] Rene Y Choi, Aaron S Coyner, Jayashree Kalpathy-Cramer, Michael F Chiang, and J Peter Campbell. Introduction to machine learning, neural networks, and deep learning. *Translational Vision Science & Technology*, 9(2):14–14, 2020.

[47] JK Rahul Jayawardana and T Sameera Bandaranayake. Analysis of optimizing neural networks and artificial intelligent models for guidance, control, and navigation systems. *International Research Journal of Modernization in Engineering, Technology and Science*, 3(3):743–759, 2021.

[48] Michael A Nielsen. *Neural networks and deep learning*, volume 25. Determination press San Francisco, CA, USA, 2015.

[49] Jason Brownlee. What is the difference between a batch and an epoch in a neural network. *Machine Learning Mastery*, 20, 2018.

[50] Keras api reference (website). `https://keras.io/about/`. Accessed: 2022-05-09.

[51] Andrey Morozov, V Solovov, R Martins, F Neves, V Domingos, and V Chepel. Ants2 package: simulation and experimental data processing for anger camera type detectors. *Journal of Instrumentation*, 11(04):P04022, 2016.

[52] G Ambrosi, M Ambrosio, C Aramo, B Bertucci, E Bissaldi, M Bitossi, A Boiano, C Bonavolontá, M Caprai, L Consiglio, et al. Characterization of fbk nuv-hd sipms for the psct camera proposed for the cherenkov telescope array. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 936:542–544, 2019.

[53] Optical grease bc-630 saint-gobain (website). `https://www.crystals.saint-gobain.com/radiation-detection-scintillators/assembly-materials`. Accessed: 2022-05-24.

[54] Vikuiti™ enhanced specular reflector (esr) (website). `https://lphe.epfl.ch/bay/www/scifimats/matprod/Materials/ESR_Usage_Guidelines.pdf`. Accessed: 2022-05-24.

[55] Matlab documentation (website). `https://www.mathworks.com/help/matlab/`. Accessed: 2022-05-29.

[56] Spyder documentation (website). `https://docs.spyder-ide.org/current/index.html`. Accessed: 2022-05-29.

[57] Suhap Sahin, Yasar Becerikli, and Suleyman Yazici. Neural network implementation in hardware using fpgas. In *International Conference on Neural Information Processing*, pages 1105–1112. Springer, 2006.

[58] Stefan Seifert, Herman T van Dam, Jan Huizenga, Ruud Vinke, Peter Dendooven, Herbert Löhner, and Dennis R Schaart. Monolithic labr3: Ce crystals on silicon photomultiplier arrays for time-of-flight positron emission tomography. *Physics in Medicine & Biology*, 57(8):2219, 2012.

[59] Giacomo Borghi, Valerio Tabacchini, and Dennis R Schaart. Towards monolithic scintillator based tof-pet systems: practical methods for detector calibration and operation. *Physics in Medicine & Biology*, 61(13):4904, 2016.

[60] SE Brunner, L Gruber, J Marton, K Suzuki, and A Hirtl. Studies on the cherenkov effect for improved time resolution of tof-pet. *IEEE Transactions on Nuclear Science*, 61(1):443–447, 2013.

Abstract in lingua italiana

I detector monolitici per i sistemi di imaging PET e TOF-PET stanno guadagnando sempre più interesse grazie alla loro maggiore sensibilità, risoluzione temporale e spaziale. Inoltre, la propagazione spaziale della luce di scintillazione può essere sfruttata per ottenere ulteriori informazioni sulla dinamica temporale della rilevazione. Questa ricerca mira a fornire un algoritmo completo di elaborazione dei dati che fornisca una predizione sulla posizione 2D dell'interazione gamma e migliori la risoluzione temporale del rivelatore. Sono state implementate due reti neurali artificiali (ANN): una ANN di posizionamento prevede la posizione dell'evento di scintillazione e una ANN temporale fornisce una correzione sulle misure del tempo di volo (TOF). La geometria del rivelatore in esame consiste in un cristallo monolitico di dimensioni 50.8x50.8x20 mm^3 accoppiato a una matrice 8x8 di SiPM. Il toolkit di simulazione ANTS2 viene utilizzato per raccogliere dati da diverse configurazioni di detector, utilizzando cristalli di $LaBr_3$(Ce) e LYSO(Ce) combinati con diversi rivestimenti. Lo studio comprende il confronto tra training dataset con diversa risoluzione spaziale, raccolti spostando una sorgente linearmente collimata a passi di 1 mm e 5 mm. I dati spaziali e temporali sono stati analizzati per usarli al meglio con le ANN. La ANN spaziale ha mostrato una FWHM spaziale di 1.805 mm sulla configurazione LYSO(Ce) + ESR. La correzione prevista dalla ANN temporale è stata in grado di ridurre l'incertezza sui TOF dal valore iniziale di 467 ps FWHM a quello corretto di 192 ps FWHM per la configurazione $LaBr_3$(Ce) + Teflon.

Parole chiave: Tomografia a emissione di positroni, TOF-PET, scintillatore monolitico, apprendimento automatico, rete neurale artificiale

Acknowledgements

First of all, I would like to express my sincere gratitude to my advisor, Professor Carlo Ettore Fiorini, who gave me the opportunity to work and learn under his guidance, this thesis was the perfect conclusion to my master's degree program, and I couldn't have done it in a better lab, it has been a tough but fulfilling journey.
I have to address a huge thank you to my co-advisor Ilenia D'Adda, without her constant help, this work would not be what it is, from start to finish she has been my point of reference, I thank her for all the time and patience she dedicated to me.
Thanks also go to Luca Buonanno, who helped me start this work, and gave me the fundamental starting ideas that led to the results I achieved. Special thanks to everyone at RadLab who in one way or another helped me in the lab.
Finally, I thank my university, Politecnico di Milano. It has been a relationship of joys and sorrows, but the best things come from the hardest challenges.