Deep Learning reconstruction with uncertainty estimation for γ photon interaction in fast scintillator detectors

Geoffrey Daniel, Mohamed-Bahi Yahiaoui, Claude Comtat, Sebastien Jan, Olga Kochebina, Jean-Marc Martinez, Viktoriya Sergeyeva, Viatcheslav Sharyy, Chi-Hsun Sung, Dominique Yvon

Engineering Applications of Artificial Intelligence 131 (2024) 107876

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Deep learning reconstruction with uncertainty estimation for γ photon interaction in fast scintillator detectors

G. Daniel [a,*], M.-B. Yahiaoui [b,a], C. Comtat [b], S. Jan [b], O. Kochebina [b], J.-M. Martinez [a], V. Sergeyeva [b,a], V. Sharyy [c,b], C.-H. Sung [c], D. Yvon [c,b]

[a] Université Paris-Saclay, CEA, Service de Génie Logiciel pour la Simulation, 91191, Gif-sur-Yvette, France
[b] Université Paris-Saclay, Inserm, CNRS, CEA, Laboratoire d'Imagerie Biomédicale Multimodale (BioMaps), 91401, Orsay, France
[c] Université Paris-Saclay, CEA, IRFU, Département de Physique des Particules, 91191, Gif-sur-Yvette, France

ARTICLE INFO

Keywords:
Deep learning
Neural networks
Uncertainty quantification
Event reconstruction algorithms
Gamma detector
PET imaging

ABSTRACT

This article presents a physics-informed deep learning method for the quantitative estimation of the spatial coordinates of gamma interactions within a monolithic scintillator, with a focus on Positron Emission Tomography (PET) imaging. A Density Neural Network approach is designed to estimate the 2-dimensional gamma photon interaction coordinates in a fast lead tungstate ($PbWO_4$) monolithic scintillator detector. We introduce a custom loss function to estimate the inherent uncertainties associated with the reconstruction process and to incorporate the physical constraints of the detector.

This unique combination allows for more robust and reliable position estimations and the obtained results demonstrate the effectiveness of the proposed approach and highlights the significant benefits of the uncertainties estimation. We discuss its potential impact on improving PET imaging quality and show how the results can be used to improve the exploitation of the model, to bring benefits to the application and how to evaluate the validity of the given prediction and the associated uncertainties. Importantly, our proposed methodology extends beyond this specific use case, as it can be generalized to other applications beyond PET imaging.

1. Introduction

1.1. Context and objectives

Gamma photon detection is used in numerous industrial, medical and security applications. It is often based on scintillating crystals coupled with a light collection and readout system. The scintillator is either pixelated, consisting on an array of small individual crystals, or continuous, made of a large monolithic block. The choice between both technologies is often based on a trade-off between sensitivity and resolution performances. For pixelated detectors, the spatial localization of the detection is provided by the crystal impacted by the gamma photon, whereas for continuous detectors, designated algorithms shall be used to derive spatial information. Several algorithms have been proposed, based either on prior knowledge of the physics of detection or on a machine learning approach (Kawula et al., 2021). The main objective of this work is the development of a physics informed deep learning method for a quantitative estimation of the spatial coordinates of the gamma interaction within a monolithic scintillator, including uncertainties. The application framework is nuclear medicine imaging, more specifically the detection of 511 keV gamma photons in Positron Emission Tomography (PET).

PET imaging is a powerful *in vivo* functional imaging modality mainly used in oncology, neurology and cardiology. It is based on the administration to the patient of a biomarker labeled with a radionuclide that decays through positron emission, followed by the detection in coincidence, outside the body, of pairs of 511 keV gamma photons resulting from the annihilation of the emitted positrons with electrons of surrounded media. The data acquisition process is followed by the tomographic reconstruction of a three-dimensional image of the biomarker distribution within the body. The quality of the reconstructed image highly depends on the performances of the gamma photon detectors in terms of sensitivity, spatial resolution, and temporal resolution. The detector spatial resolution has a direct impact on the contrast recovery of small structures in the image. A higher detection efficiency translates into a higher number of detected coincidences, resulting in a better signal-to-noise ratio (SNR) in the PET image. The improvement of the coincidence resolving time (CRT), characterizing the capability of a pair of detectors to resolve the difference between

* Corresponding author.
E-mail address: geoffrey.daniel@cea.fr (G. Daniel).

https://doi.org/10.1016/j.engappai.2024.107876
Received 1 June 2023; Received in revised form 1 December 2023; Accepted 8 January 2024
Available online 13 January 2024

the times of interaction of the two 511 keV gamma photons detected in coincidence, also helps for increasing the SNR in the image (Schaart, 2021). This principle is referred to as time-of-flight (ToF) PET.

State of the art clinical PET systems are based on pixelated detectors made of LSO (lutetium oxyorthosilicate) or LYSO (lutetium–yttrium oxyorthosilicate) crystals, with a pixel pitch between 3.5 and 5 mm. These PET systems have, at best, a CRT of 210 ps (van Sluis et al., 2019), an intrinsic spatial resolution in the reconstructed image of 3.5 mm (van Sluis et al., 2019) (about 3 mm at the detector level), and an absolute sensitivity of $\sim$ 20 counts/sec/kBq (i.e. $\sim$ 2%) (Grant et al., 2016). Major efforts are being made worldwide in nuclear instrumentation research groups to improve these parameters, in particular to decrease the CRT below 100 ps, ideally down to 10 ps (Lecoq et al., 2020). There is a trend toward the development of PET detectors based on monolithic crystals for their higher sensitivity, since there are no intercrystal gaps (Gonzalez-Montoro et al., 2021). Based on the scintillation light distribution readout, dedicated neural networks have recently been proposed to provide a single 3-dimensional gamma photon interaction position within the monolithic crystal (see, for example (Belov et al., 2023; Freire et al., 2022; Carra et al., 2022; Jaliparthi et al., 2021; Kawula et al., 2021)). All these studies demonstrate the benefit of using a machine learning approach for an accurate position estimation.

In this study, we propose to address the question of the reconstruction of the 2-dimensional gamma photon interaction coordinates for a fast lead tungstate ($PbWO_4$) monolithic scintillator detector (Yvon et al., 2020), that is not straightforward due to the specificity of our acquisition system which is not a conventional pixelated photodetector. Our system consists indeed in the use of a micro-channel plate photomultiplier tube (MCP-PMT) which necessitates a dedicated process to reconstruct gamma photon interaction parameters from the acquired signal, as we describe in this paper. Moreover, we aim not only to perform the reconstruction but to associate uncertainties on the reconstruction. We will exploit and validate these uncertainties to assess the reliability of the network prediction. This approach is new in this field and the methodology we propose can be extended to other applications on signal processing or sensor data analysis. In the future, these spatial coordinates uncertainties could be used during the tomographic reconstruction and potentially improve the quality of the PET image.

1.2. Contributions and organization of the paper

Through this paper, we present a specific methodology that relies on the paradigm of uncertainty estimation, which is a growing topic in the scientific literature on Artificial Intelligence. Several approaches are usually proposed, such as Bayesian Neural Network or Deep Ensembles (Gawlikowski et al., 2021). In this paper, we use a Density Neural Network, which is a relevant approach when the uncertainties are dominated by fluctuations and random phenomenon in the data, such as the source of variability during the scintillation process in our detectors. This approach brings additional information thanks to the uncertainties that are used to gain information on the prediction of neural networks.

A dedicated methodology of exploitation and validation is presented in this paper:

- The design of a dedicated loss function, based on a negative log-likelihood of an assumed distribution. This distribution is often assumed as a Gaussian distribution, but it is not restricted to this type of distribution and can be adapted to the problem to address. In our case, we decide to add physical constraints through a truncation of a Gaussian distribution according to the detector edges, which requires a specific implementation of the loss function. We show that this approach brings improvements for reconstructing specific events, especially close to the detector.
- The estimated uncertainties can be exploited to complete the prediction of the neural network. The uncertainty can be evaluated event by event to assess the confidence on the prediction of the model. In our case, we show that we can statistically improve the reconstruction resolution and this feature will be helpful in the exploitation of the detector edges.
- The uncertainties are self-estimated by the Neural Network, thus they are prone to the same biases as the classical prediction. Consequently, the validation of the estimated uncertainty is required. We use coverage plots as a possible approach to empirically verify the relevance and reliability of these uncertainties.

We want to highlight that this whole methodology can be beneficial for many applications of Deep Neural Network in regression tasks, beyond the processing of scintillation detectors signals.

The paper is organized as follows. In Section 2, we describe the gamma photon detector based on $PbWO_4$, the Monte Carlo simulation used to generate the training and testing datasets, and the preprocessing of the detector raw data. Section 3 focuses on the methodology of the deep learning approach used for this study, including the description of a baseline method for performance comparisons and the definition and the architecture of the selected Density Neural Network. Results are presented in Section 4. These results and the methodology are discussed in Section 5.

2. Materials

2.1. Detector description

The ClearMind gamma detector (Fig. 1) is composed of a MCP-PMT sealed by a monolithic $PbWO_4$ crystal, acting both as the gamma conversion crystal and as the optical window of the MCP-PMT. A high quantum efficiency photoelectric layer is deposited on its inner face. The direct deposition of a photocathode with a refraction index superior to the refraction index of the $PbWO_4$ crystal allows us to avoid total reflection at the crystal/photocathode interface, thus maximizing the photon collection efficiency of the module (Yvon et al., 2020). The use of this ”scintronic” crystal as an entrance window of a MCP-PMT makes it possible to optimize the time resolution thanks to the excellent electron transit time spread (~60 ps FWHM) to the detection anodes provided by this type of photodetector.

The $PbWO_4$ crystal, homogeneously doped, has a surface of 59 mm$\times$ 59 mm a thickness of 5 mm, and is provided by CRYTUR (CRYTUR, spol. s r.o., Na Lukách 2283 511 01 Turnov Czech Republic, 2020). The photocathode deposit and the integration of the device into a MCP-PMT structure is handled by the PHOTEK company, based on its MAPMT-253 design (MAPMT-253 Multi-Anode MCP-P.M.T. Datasheet, from Photek Inc., 2020). We developed a signal readout system for this device using 32 transmission lines as shown in Fig. 1, Follin et al. (2022). The signals are read out at both ends of the transmission lines, amplified and then sampled by a SAMPIC WaveShape Recorder (Breton et al., 2020).

Typically a 511 keV energy deposit in the crystal produces 185 optical photons mostly isotropically. Out of these, $\sim$ 20% are collected by the 53 mm$\times$53 mm photocathode and generate detected photoelectrons that are collected and amplified by the MCP-PMT photodetector.

Many processes in the signal formation involves random features. For example, the photon direction and time of production, the photoelectron production probability, the gain of the Micro Channel Plate used for electron multiplication, the time transit of the electron propagation through the MCP-PMT, the noise of readout amplifiers are best described using parameterized random variables. These are necessary to describe the pulses shapes produced by a single photoelectron (SPE).

Each SPE induces a signal on typically three readout lines. Thus the typical 30 SPE signals pileup at the output of the transmission lines (Fig. 2), and build the event signal registered by the SAMPIC module.

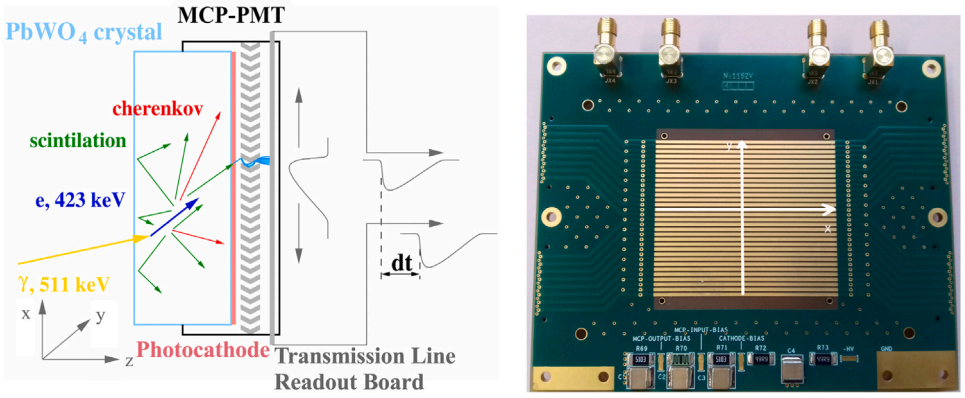

Fig. 1. Left: Schematic diagram of the ClearMind detection module. A 511 keV gamma-ray interaction in the crystal produces scintillation and Cherenkov photons that are converted by the photocathode to photoelectrons. These photoelectrons are then multiplied by the MCP-PMT and induce signals on the transmission lines (TLs). Signals from the left and right ends of each TL are amplified by 40 dB amplifiers and digitized by a SAMPIC module. **Right:** Transmission lines Printed Circuit Board (PCB). The axis x and y corresponds to the coordinate system that we use to locate the interaction position.

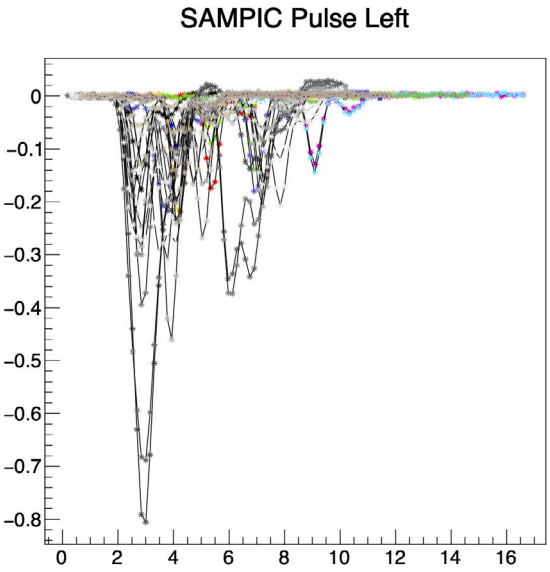

Fig. 2. Set of pulses as registered by the SAMPIC waveshape recorder for a 511 keV energy deposit. For clarity purpose, only the pulses registered on one side of the transmission lines are shown (half of the set).

2.2. Simulation

In order to create a sufficiently large and unbiased database to train and test our reconstruction algorithms, we developed a detailed simulation of the ClearMind detector (Sung et al., 2023). The knowledge of the hidden physical processes available in Monte Carlo simulations is necessary to provide the ground truth (target) for the training of the Machine Learning models. It is also very convenient to assess the intrinsic performances of our algorithms.

This simulation is based on the Gate v9.0 (Jan et al., 2004; Sarrut et al., 2021-05) / Geant4 v7.0 (Agostinelli et al., 2003; Allison et al., 2006, 2016) software allowing to simulate in full details the interaction of the particle with matter and optical photons generation and tracking. Furthermore, we have developed specialized software to simulate the photodetector, analog, and digital electronic components. Necessary parameters have been extracted from dedicated measurements (Follin et al., 2021, 2022). This simulation includes the following main parts of the detector response.

1. The gamma interaction in the crystal accounts for three processes: photoelectric conversion, Compton scattering and Rayleigh diffusion. The two first processes produce relativistic electron that emits visible photons through two mechanisms: Cherenkov radiation and scintillation (~20 and ~165 photons for 511 keV γ-quanta respectively).
2. Each optical photon is propagated individually by the simulation program. During the propagation all main physical effects are taken into account: photon absorption inside the $PbWO_4$ crystal, reflection or absorption on the crystal borders for the different types of the crystal surface (polished, ground, absorbing), escape of photons from the crystal into the air.
3. Photocathode simulation includes the Fresnel reflection of visible photons at photocathode boundaries, absorption of photons by the photocathode and extraction of generated photoelectrons as a function of the photon wavelength. As a result we compute, assuming a photocathode of nominal efficiency, that we produce in average 30 photoelectrons for a 511 keV γ-ray photoelectric conversion in the crystal and 75% of events contain at least one Cherenkov photon converted into a photoelectron.
4. We then simulate the propagation and the multiplication of individual photoelectrons generated by the photocathode in the MCP-PMT and parametrize the main PMT response features: time response, PMT gain and gain fluctuation, signal sharing between different output anodes.
5. Finally, we simulate the signal readout through the transmission lines with realistic signal shapes, taking into account the possible overlay of several photoelectrons, electronics noise and digitization sampling.

Most of the simulation parameters are adjusted to the results obtained by the characterization of the first prototype using pulsed laser in the single-photon regime. More details about the simulation could be found elsewhere (Sung, 2022; Sung et al., 2023).

2.3. Waveform preprocessing and input data shaping

Depending on the energy deposited in the crystal, the SAMPIC module records from 2 to 64 pulse shapes, typically about 30 for a deposited energy of 511 keV. This corresponds to about 10 kBytes of raw data per event. This volume of data is to be compared with the volume of data to be reconstructed, the properties of the gamma interaction in the crystal: 3D position, time, deposited energy, interaction multiplicity, typically 25 bytes per event. The acquired data are therefore redundant, but also intricate. The parameters to be reconstructed are encoded in a complex way and are mixed in the pulse shapes.

Thus reconstructing the properties of the gamma interaction in the crystal is a complex task.

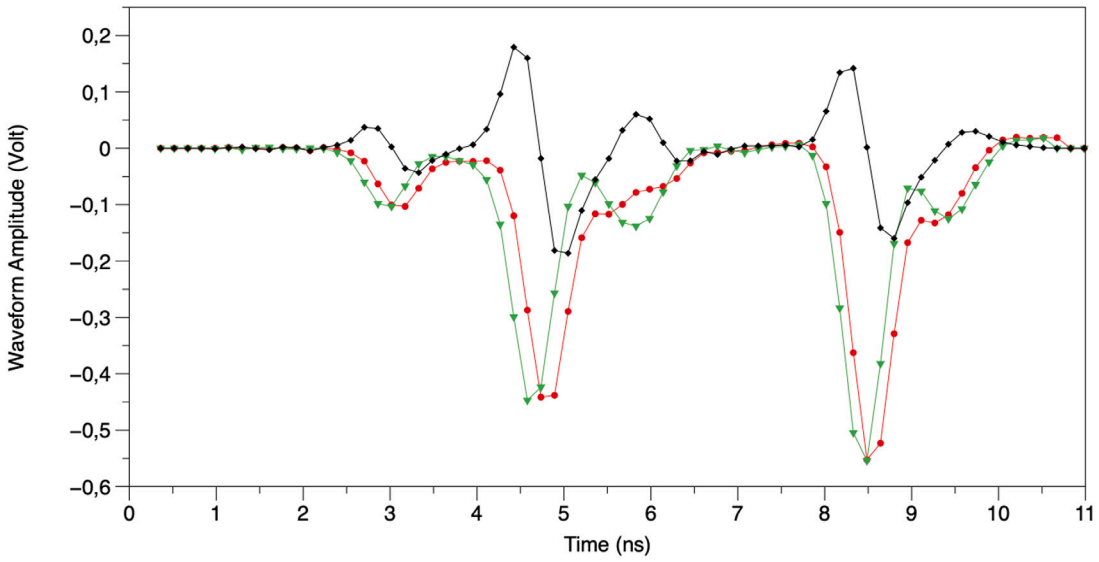

Fig. 3. Waveforms registered on one triggered line l. Red and green lines are the left $F_{l,\mathrm{Left}}(t_j)$ and right $F_{l,\mathrm{Right}}(t_j)$, time shifted, registered pulses shapes. Black line shows the time difference waveform $F_{l,\mathrm{Left}}(t_j) - F_{l,\mathrm{Right}}(t_j)$. We identify on this line three pulses clusters at 4.6 ns, 6 ns and 8.5 ns. For each of them, the time difference curve shows a bipolar shape, correlated to the position of each photoelectron charge induction along the readout line.

We first use our knowledge of the physics of the detector to calculate on the raw data a set of statistical variables, so called "parameter observables" highly correlated to the parameters of the gamma-ray interaction to be reconstructed, as well as a second set of "bias observables" which monitor the known undesirable instrumental effects (saturation of the acquisition electronics, edge effects, etc...). The data volume of the observables is 100 bytes per event. These observables are then used as inputs of shallow, fully connected neural networks, whose training takes only a few minutes. The development cycle is fast, at the cost of a loss of information that is difficult to anticipate.

In the following paragraph, we explain the parameter observables developed to correlate to the features targeted by the neural networks presented in this paper: the gamma interaction position x_{pos} (*position along the transmission lines*) and y_{pos} (*perpendicular to the transmission lines*).

First, for each transmission line l, that triggered the SAMPIC acquisition, the digitized pulse shapes are acquired: $F_{l,\mathrm{Left}}(t_j)$ and $F_{l,\mathrm{Right}}(t_j)$ at times t_j for the Left and Right transmission line, where the index j corresponds to the sampling time. Examples of acquired pulse shapes are shown in Fig. 3.

We first calculate the electric charges collected (integral of the pulse shape current over time) at both the end of the 32 transmission lines C_l.

We compute two interaction observables expected to be correlated to y_{pos} :

- We select the line with the largest collected charge and its two neighbors. We fit a parabola on these three values. The position of the maximum of this parabola is the first observable.
- The second observable is the median of the distribution of line numbers weighted by the charge collected on these lines : $\mathrm{Med}_L = \mathrm{median}(l, C_l)$. It is common that one line carries a large fraction of the total charge collected over the 32 readout lines registered in an event. In order to extract as much information as possible from the surrounding line charge values, we developed an "upgraded" median algorithm documented in Appendix B. This is the "median" algorithm we will use for all the observables calculations.

Quantifying observables expected to correlate to x_{pos} is more complex. It turns out that on some lines, we can identify few pulses separated in time in the recorded pulse shapes. We denote each identified pulse by the index p. When this happens, we quantify the pulse detection times ($T_{l,p,\mathrm{Left}}$ and $T_{l,p,\mathrm{Right}}$) and the total charge $C_{l,p}$ of each of these pulses. The time difference $\Delta t_{l,p} = T_{l,p,\mathrm{Left}} - T_{l,p,\mathrm{Right}}$ is correlated to the position of the pulse induced by this photoelectron along this line. We then calculate the three following observables:

- the *median* of the $\mathrm{Med}_{\Delta t} = \mathrm{median}(\Delta t_{l,p}, C_{l,p})$ distribution (all lines and pulses), weighted by the integrated charge of the each pulses. The details of the algorithm for the weighted median are given in Appendix B;
- the *mean* of the same distribution $\mathrm{Mean}_{\Delta t} = \mathrm{mean}(\Delta t_{l,p}, C_{l,p})$;
- Δt_{lmax}, the time difference Δt of the largest pulse of the line (greatest charge collected).

We also used the shape of the digitized pulses $F_{l,p,\mathrm{Left}}(t_j)$ and $F_{l,p,\mathrm{Right}}(t_j)$. The $\mathrm{Diff}_{l,p}(t_j) = F_{l,p,\mathrm{Left}}(t_j) - F_{l,p,\mathrm{Right}}(t_j)$ show bipolar shapes depending on the position of the pulse injection along the line. We integrated the **F**irst and **S**econd components of the bipolar shape $\mathrm{Int}_{l,p,F}$ and $\mathrm{Int}_{l,p,S}$. We then compute the $\mathrm{Bipol}_{l,p} = (\mathrm{Int}_{l,p,F} - \mathrm{Int}_{l,p,S})/C_{l,p}$, for each pulse. The saved observables are then:

- the *weighted median* of the distribution $\mathrm{Med}_{\mathrm{Bipol}} = \mathrm{median}(\mathrm{Bipol}_{l,p})$ (see Appendix B for the details)
- the *mean* of the distribution $\mathrm{Mean}_{\mathrm{Bipol}} = \mathrm{mean}(\mathrm{Bipol}_{l,p})$.

Along with these observables, we also compute parameters and bias observables relevant for reconstruction the other properties gamma interactions in the crystal (see Appendix A). Some of them are also expected to be relevant for the uncertainty estimation. These **23 observables are the inputs of the following neural networks**.

An alternative approach would be to use the full waveforms' data as inputs for convolutional deep neural networks. However, preliminary tests showed that the complexity of the data required to implement deeper network structures, involving training times of days on our GPU card, thus with slower development cycles. We are considering this option for future works.

3. Methodology

3.1. Baseline method

The simplest approach to gamma-interaction reconstruction is using the fact that the highest density of the detected photons corresponds

to the coordinates of the interaction point. This suggests to use the transmission line with the maximum detected charge as the reference line. To reconstruct the coordinate across lines, y_R, we calculate the weighted average of center-of-line for this and two neighboring lines:

$$y_R = \frac{\sum_{l=i-1}^{i+1} y_l C_l}{\sum_{l=i-1}^{i+1} C_l}\,, \tag{1}$$

where y_l is the y-coordinate of the line center, C_l is the charge of line l (only the negative signal part is used for the charge calculation), i is the reference line number.

The coordinate along the lines, x_R is reconstructed as

$$x_R = \frac{(t_{\text{Right}} - t_{\text{Left}})}{2} \times v_{\text{signal}}\,, \tag{2}$$

where t_{Right} and t_{Left} are the time measured at the right and left ends of line i respectively and v_{signal} is the signal propagation speed, assumed to be 35% of speed of light in the simulation.

This approach, based only on expert knowledge of the physical processes is used as a reference for comparison with the performance of our approach based on neural networks.

3.2. ML approach

This section describes our approach based on a supervised Machine Learning algorithm, that takes as input the 23 preprocessed variables described in Section 2.3 and Appendix A in order to predict the interaction position. We use a Neural Network model that is able to provide both the prediction and an estimation of an uncertainty associated to this prediction thanks to the paradigm of Density Neural Networks. On the contrary to the baseline method, the use of Neural Network is mainly based on the exploitation of simulated data to build a model that link the observable quantities to the reconstruction of the gamma interaction position. Moreover, the baseline method is not able to provide uncertainties associated to the reconstruction, which is a key difference with our approach.

3.2.1. Density neural network

The conventional approach to perform a regression through a Neural Network is to define a loss function that measures the discrepancy between the predictions of the neural network and the expected values. With classical notations, let $D = \{(x_i, y_i), i = 1, 2, \dots, N\}$ be the database and $\mu_\theta(x)$ the prediction of the expected value of the neural network parameterized by θ. By noting $\|\bullet\|$ the Euclidean norm, a usual loss function is the Mean Square Error :

$$\text{MSE}(\theta) = \frac{1}{N}\sum_{i=1}^{N} \|y_i - \mu_\theta(x_i)\|^2. \tag{3}$$

The learning procedure consists in finding parameters θ^*, among the space of possible parameters Θ, that minimizes the Mean Squared Error. It corresponds to solving the following problem:

$$\theta^* = \underset{\theta\in\Theta}{\text{argmin}}\ \text{MSE}(\theta). \tag{4}$$

The Density Neural Network approach proposes to make the assumption that the expected outputs y_i are a realization of a probability distribution such as the normal distribution $y_i \sim \mathcal{N}(\mu_\theta(x_i), \sigma^2_\theta(x_i))$ or a mixture of normal distributions (Bishop, 1994). However, this method can be adapted to other distributions and we propose to use a distribution that considers some specific physical constraints of our application.

The main physical constraint on the interaction position is that this interaction must be located in the limits of the detector. Therefore, we decide to use a truncated normal distribution in order to insure this constraint.

Beforehand in order to avoid confusion with (x, y) coordinates of the position of the interaction in the detector, we will note s the inputs data from the preprocessing describes in Section 2.3. So formally, our neural network is now expected to output four values for each inputs s_i, the two means $\mu_{\theta,x}(\mathrm{s_i}), \mu_{\theta,y}(\mathrm{s_i})$ and the two scale parameters $\sigma^2_{\theta,x}(\mathrm{s_i}), \sigma^2_{\theta,y}(\mathrm{s_i})$ of the assumed truncated Gaussian law. We have assumed independent normal laws on x_i and y_i. In this case, the hypothesis that each output x_i, y_i follows a truncated normal distribution in $[a, b]$ leads to the two following distributions relative to the two coordinates with $z = x$ or $z = y$:

$$p(z_i|\mu_{\theta,z}(\mathbf{s_i}), \sigma^2_{\theta,z}(\mathbf{s_i})) = \underbrace{\frac{1}{\sqrt{2\pi\sigma^2_{\theta,z}(\mathbf{s_i})}} \exp\left(-\frac{(z_i - \mu_{\theta,z}(\mathbf{s_i}))^2}{2\sigma^2_{\theta,z}(\mathbf{s_i})}\right)}_{\text{Classical Gaussian Likelihood}} \times \underbrace{\frac{1}{\boldsymbol{\Phi}\left(\frac{b-\mu_{\theta,z}(\mathbf{s_i})}{\sigma_{\theta,z}(\mathbf{s_i})}\right) - \boldsymbol{\Phi}\left(\frac{a-\mu_{\theta,z}(\mathbf{s_i})}{\sigma_{\theta,z}(\mathbf{s_i})}\right)}}_{\text{With truncation in } [a,b]} \tag{5}$$

where $\boldsymbol{\Phi}$ is the Cumulative Distribution Function of the standard normal distribution, a and b are the truncation boundary of the truncated normal distribution. In our application, a and b represent the limits of the detector: $a = -30$ mm and $b = 30$ mm. In our case, the condition $a \le (x_i, y_i) \le b$ is assumed to be always verified : an interaction occurs inside the detector; otherwise it is a simulation error and we should not consider it. So, by using Eq. (5), we derive the likelihood of the parameters on the whole dataset:

$$L(\theta) = \prod_i \prod_{z\in\{x,y\}} p(z_i|\mu_{\theta,z}(\mathbf{s_i}), \sigma^2_{\theta,z}(\mathbf{s_i})) \tag{6}$$

Maximizing this likelihood is equivalent to minimizing the negative log-likelihood, that is easier to handle, which leads to the following loss function:

$$\begin{aligned} l(\theta) &= -\log(L(\theta)) \\ &= \sum_i \sum_{z\in\{x,y\}} \log\left(\boldsymbol{\Phi}\left(\frac{b-\mu_{\theta,z}(\mathbf{s_i})}{\sigma_{\theta,z}(\mathbf{s_i})}\right) - \boldsymbol{\Phi}\left(\frac{a-\mu_{\theta,z}(\mathbf{s_i})}{\sigma_{\theta,z}(\mathbf{s_i})}\right)\right) \\ &+ \frac{1}{2}\log(2\pi\sigma^2_\theta(\mathbf{s_i})) + \frac{(z_i - \mu_{\theta,z}(\mathbf{s_i}))^2}{\sigma^2_{\theta,z}(\mathbf{s_i})} \end{aligned} \tag{7}$$

We want to highlight that the conventional problem with MSE defined in Eq. (3) is a special case of Eq. (7) and $a = -\infty$, $b = +\infty$ (no truncature) and where the scale parameters $\sigma^2_{\theta,x}(\mathbf{s_i})$ and $\sigma^2_{\theta,y}(\mathbf{s_i})$ are assumed to be constant (homoscedasticity hypothesis). Finally, the optimization problem consists now in finding the parameters θ^* that minimize this new loss function:

$$\theta^* = \underset{\theta\in\Theta}{\text{argmin}}\ l(\theta) \tag{8}$$

We can solve this problem by using a conventional gradient descent optimization in order to train the neural network and find optimal parameters.

This approach brings two main components:

- the use of a Density Neural Network allows the estimation of an uncertainty associated to the network output through the prediction of the terms $\sigma^2_{\theta,x}(\mathbf{s_i})$ and $\sigma^2_{\theta,y}(\mathbf{s_i})$. These variances are specific to each example and represent a higher or lower uncertainty on the predicted positions $\mu_{\theta,x}(\mathbf{s_i})$ or $\mu_{\theta,y}(\mathbf{s_i})$ according to the inputs $\mathbf{s_i}$ of the neural network. For instance, in our application, we can expect an interaction closer to the edge of the detector is less precisely located than an interaction at the center of the detector. We highlight that this uncertainty is self-estimated by the neural network and consequently must be empirically validated, we bring elements for this validation in Section 4.3. This uncertainty can be seen as aleatoric uncertainty (Kendall and Gal, 2017) due to the randomness of the phenomenons involved during the detection and measurements (Cherenkov and scintillation production, loss of optical photons, randomness in the pulse shape creation...).

- the use of a truncated normal hypothesis allows us to consider the physical constraints. In Section 4, we show that this property is essential to obtain better performances for interaction close to the edges of the detector, compared to the use of the classical MSE approach.

3.2.2. Architecture and training

We develop two neural networks: one neural network associated to the conventional Mean Squared Error loss function and one neural network associated to our custom loss function with the truncated normal hypothesis. For both architectures, all of the hidden layers are similar: we use six hidden fully-connected layers with 256 neurons each and a tanh activation function. The difference is the output layer:

- the output layer for the neural network associated to the MSE loss has two neurons (one for each coordinate), with a scaled tanh function: $\frac{b-a}{2}\tanh$ in order to insure that the prediction is located inside the detector.
- the output layer for the neural network associated to the truncated normal loss function has four output neurons. The first ones output $\mu_{\theta,x}(\mathbf{s_i})$ and $\mu_{\theta,y}(\mathbf{s_i})$ and the activation function is the previous scaled tanh function. The second ones output $\sigma_{\theta,x}(\mathbf{s_i})$ and $\sigma_{\theta,x}(\mathbf{s_i})$. The activation function is softplus$(\alpha)+\epsilon$ where α is the output before applying the activation function. ϵ is a small constant that ensures that $\sigma_{\theta,x}(\mathbf{s_i})$ and $\sigma_{\theta,y}(\mathbf{s_i})$ are not too close to 0 and avoids divergence issues. In our case, we choose $\epsilon = 10^{-6}$ mm.

The neural networks and their trainings are computed by using the Tensorflow (Abadi et al., 2015) library. The optimizer is the Adam algorithm. The training is stopped by an early stopping procedure that monitors the validation loss; we use randomly 20% of our data as validation data. The code has been executed on a NVIDIA RTX Quadro 5000. The training time is a dozen of minutes.

3.2.3. Generated datasets

We used three datasets to train and test our supervised Machine learning algorithms. All of them use the detector modeling explained in Section 2, corresponding to a PbWO4 scintillation crystal of 59 mm × 59 mm. The inputs are the 23 observables described in Section 2.3 and Appendix A from the 32 transmission lines and the outputs are both coordinate X and Y of the gamma interaction, which are saved thanks to the Geant4 simulation.

1. For the training set, we simulate a Gamma photon source shaped in a 6 cm cube. The energy spectrum has been adjusted to increase the probability of high energy deposits in the $PbWO_4$ and thus generate an approximately flat deposited energy spectrum. Thus we generate 7 times more 1.2 MeV than 300 keV photons. The gamma rays impinge the $PbWO_4$ crystal perpendicularly and uniformly over the entire surface of the optical window. This training batch contains 450 000 events.
2. The first test dataset simulates an grid of 9 × 9 511-keV gamma ray point sources, regularly spaced by 7 mm. Again the gamma rays impinge the $PbWO_4$ crystal perpendicularly to the surface of the optical window. This test dataset contains 300 000 events.
3. The second test dataset simulates a 6 cm cube-shaped Gamma photon source, mono-energetic at 511 keV. The gamma rays impinge the $PbWO_4$ crystal perpendicularly and uniformly on the whole surface of the optical window. This test dataset contains 600 000 events. We design this test dataset with a high number of examples in order to conduct an accurate analysis of the performances of the neural network considering the uncertainty estimation.

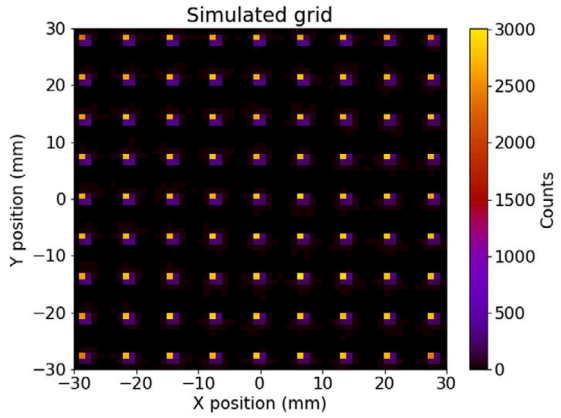

Fig. 4. Simulated sources — Expected positions.

4. Results

4.1. Evaluation on the grid of sources — First test dataset

The positions of the grid of sources are represented on Fig. 4. This configuration helps visualizing directly some properties of the different reconstruction algorithms. The results are presented on Fig. 5.

They show that the algorithms are all able to reconstruct the sources in the center area of the detector, between −20 mm and +20 mm in both directions. However, the baseline method and the conventional Neural Network associated to a MSE loss function are not able to give a prediction on the edge of the detector. These predictions are even incorrectly attributed to other position at −25 mm and +25 mm for the conventional MSE approach. This "folding" of the predictions will imply a lower confidence of the reconstruction in this area, as we show in Section 4.2. On the contrary, the use of the truncated Gaussian likelihood is able to provide positions close to the edge and avoid the "folding" effect. This property shows a first advantage of the use of this custom loss function.

The bottom right figure introduces the benefits of the uncertainty evaluation. To produce this figure, we use a weighting on each event i according to the predicted uncertainties terms $\sigma_{\theta,x}(\mathbf{s_i})$ for the X position, and $\sigma_{\theta,y}(\mathbf{s_i})$ for the Y position. The weights are computed using the following equation:

$$w_i = \frac{(\sigma_{\theta,x}(\mathbf{s_i})^2 + \sigma_{\theta,y}(\mathbf{s_i})^2)^{-1}}{F}, \tag{9}$$

where F is a normalization factor such that $\sum_i(w_i)$ is equal to the number of detected events on the considered bin of the histogram.

$$F = \sum_{i\in\text{bin}} (\sigma_{\theta,x}(\mathbf{s_i})^2 + \sigma_{\theta,y}(\mathbf{s_i})^2)^{-1} \tag{10}$$

This weighting penalizes the events with a high predicted uncertainty. The result shows a reduction of the spreading effect, making the reconstruction more accurate on the detector. We quantify this gain in accuracy in Section 4.2.

4.2. Evaluation on the uniform simulation — Second test dataset

For the global evaluation of the performances, we use the simulation of a uniform distribution of sources (cube source) in front of the detector. The direct reconstruction provided by the algorithms is shown on Fig. 6. The results show the same properties as the grid simulation, especially the inability of the baseline method and the conventional MSE to reconstruct the edges of the detector. We can see a partial "folding" effect on the reconstruction with the truncated Gaussian likelihood. This effect is attenuated by applying the weighting as defined in Eq. (9).

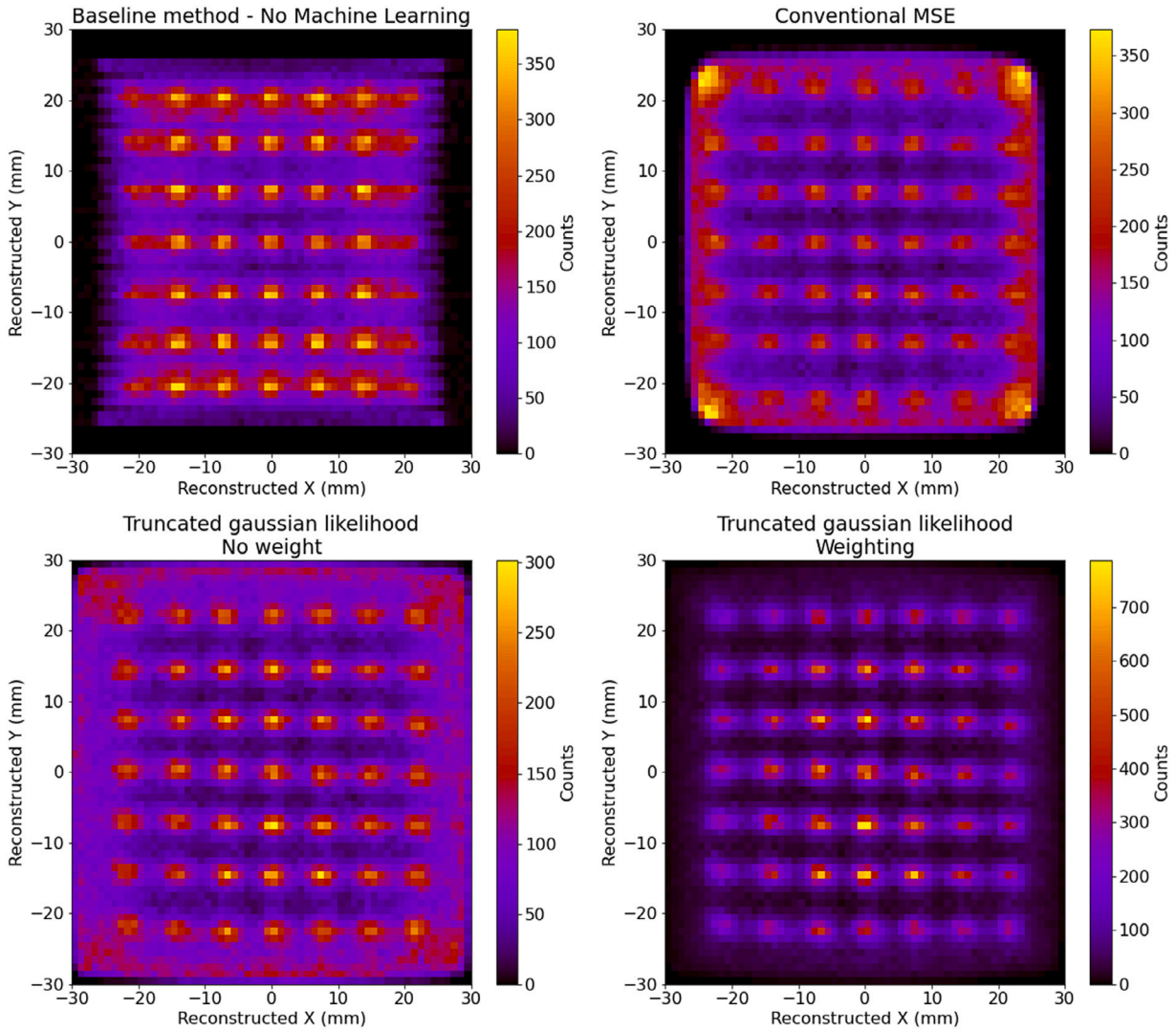

Fig. 5. Grid reconstruction by the different methods.

Fig. 7 shows for each true position the average 2D distance to the reconstructed position. A high average 2D distance in this case corresponds to a high spread of the reconstructions. At the center of the detector, the baseline method provides less spread reconstructions, between 2 and 3 mm, than the conventional MSE and our approach with the truncated Gaussian likelihood, between 3 and 4 mm. However, the area of high spread, over 6 mm, is larger for the baseline method than the machine learning method. We show later that this default highly affects the precision of the reconstruction, even at the center of the detector.

The use of the weighting helps the truncated Gaussian likelihood to give more importance to the most certain events, leading to a spread between 1 and 2 mm at the center of the detector, between 2 and 4 mm in the intermediate area. The area with the higher spread, over 6 mm, is thinner than the other algorithms.

Fig. 8 shows for each predicted position the average 2D distance to the true position. A high average 2D distance in this case corresponds to a low precision of the reconstructions. Black areas correspond to area with no predicted position due to the inability of the algorithms to reconstruct some positions, as shown on Fig. 6. The baseline method gives very low accurate results, mostly over 6 mm, even at the center of the detector. By combining with the previous information, we can give the following interpretation:

- if the algorithm provides a localization at the center of the detector, we cannot trust the algorithm since the reconstruction error is over 5 mm, as shown on Fig. 8 ;
- if the true position is the center of the detector, we can trust the algorithm because the reconstruction error is between 2 and 3 mm, as shown on Fig. 7. However, for a real application, we cannot access to the true position, then we cannot exploit this information.

The conventional MSE shows a better precision than the baseline method, with a reconstruction error between 3 mm at the center of the detector to 5 mm close to the edges. The areas with very high reconstruction error, over 6 mm, are limited. The truncated Gaussian likelihood shows similar performances as the conventional MSE for the precision, with the ability to reconstruct the edges, with less precision. The application of the weighting improves the results, providing a precision between 1 and 2 mm at the center of the detector, between 3 and 5 mm in the intermediate area and over 6 mm in a thin band close to the edge. This band corresponds to the part of the detector without optical layer and it is expected to have a degradation of the performances in this area.

Finally, Figs. 9 and 10 show the overall results, considering the whole test dataset. The histograms represent the global reconstruction errors, without selecting any position. We also compute the Root Mean Squared Error (RMSE) of the predictions.

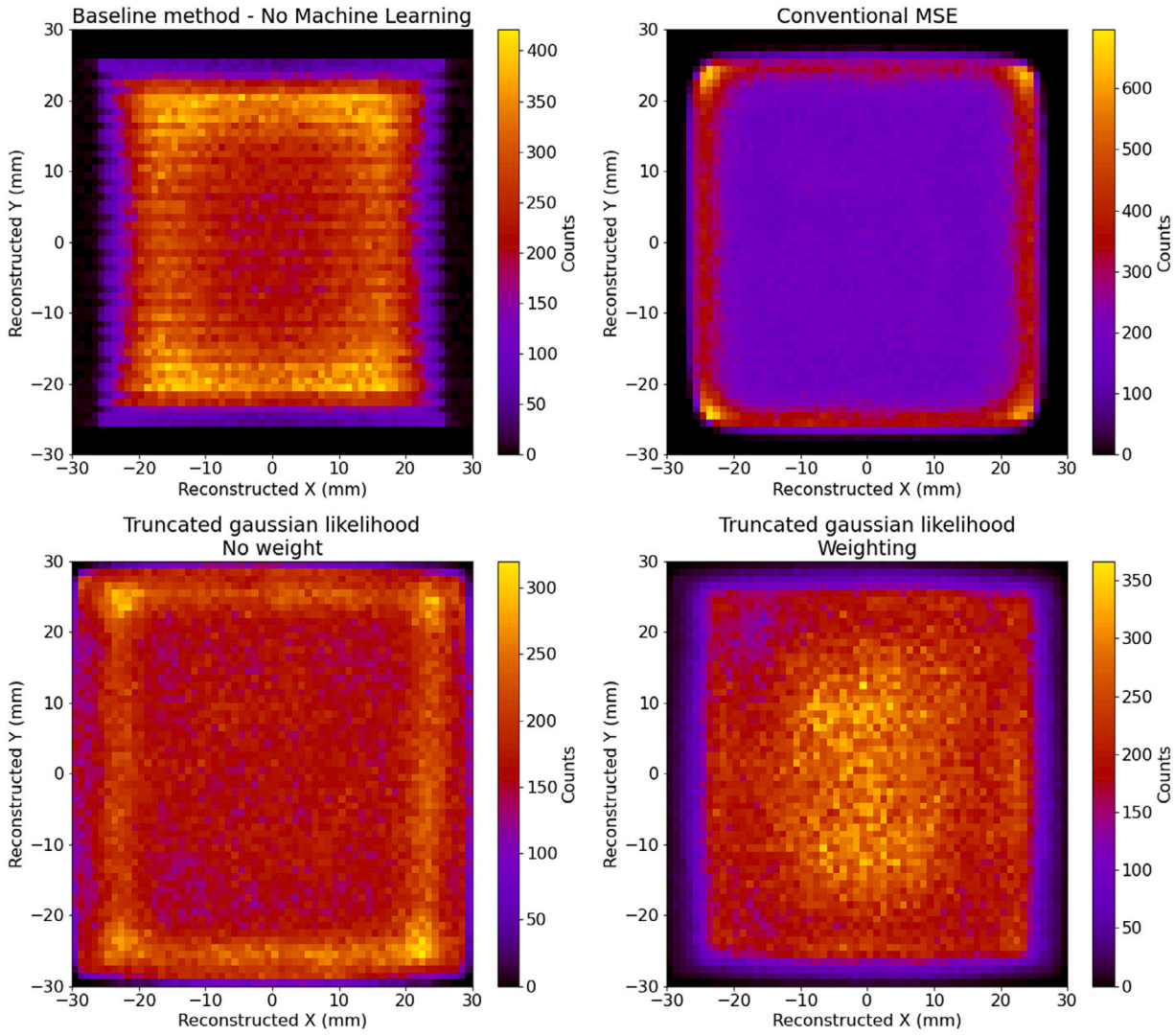

Fig. 6. Reconstruction of the uniform simulation.

Table 1
Table of performances. Best values are in bold.

Quantity	Baseline method	Conventional MSE	Truncated Gaussian lik.	Truncated Gaussian lik. weighting
Mean error X (mm)	**−0.01**	**0.01**	−0.32	−0.39
RMSE X (mm)	8.05	4.8	5.11	**2.61**
Std dev X (mm)	8.05	4.8	5.1	**2.58**
Mean error Y (mm)	**0.00**	0.19	0.04	0.02
RMSE Y (mm)	5.5	4.72	5.08	**1.91**
Std dev Y (mm)	5.5	4.72	5.08	**1.91**
Mean 2D error (mm)	6.07	4.36	4.56	**2.51**

For both coordinates, the baseline method shows the worse performances. This result shows the advantage to use a Machine Learning approach to address this reconstruction problem, due to the complex link between the observable variables and the positions.

We apply a new 1D-weighting dedicated to each coordinate:

$$w_{X,i} = \frac{(\sigma_{\theta,x}(\mathbf{s_i}))^{-2}}{\sum_j (\sigma_{\theta,x}(\mathbf{s_j}))^{-2}} \tag{11}$$

$$w_{Y,i} = \frac{(\sigma_{\theta,y}(\mathbf{s_i}))^{-2}}{\sum_j (\sigma_{\theta,y}(\mathbf{s_j}))^{-2}} \tag{12}$$

On average, the conventional MSE gives slightly better results than the truncated Gaussian likelihood in terms of RMSE. We can expect this behavior because the MSE approach is specifically designed to minimize the MSE, thus its squared root the RMSE.

Table 1 summarizes several performances such as the mean error (the bias) on the histogram, RMSE and the Standard Deviation (spread of the distribution without bias). A small bias is observed for the reconstructions with the truncated Gaussian likelihood methods for the X coordinate, under 1 mm. On the contrary, the conventional MSE presents a bias for the Y coordinate reconstruction. On the other metrics, the weighting applied to the truncated Gaussian likelihood outperforms the other methods: the application of this weighting helps improving the performances by giving less importance to uncertain reconstructions. Especially, the RMSE reaches 2.61 mm for the X coordinate and 1.91 mm for the Y coordinate.

Finally, the Mean 2D error corresponds to the mean of the Euclidean distances between the predicted position and the expected position. Regarding this metric, the Machine Learning algorithms show better

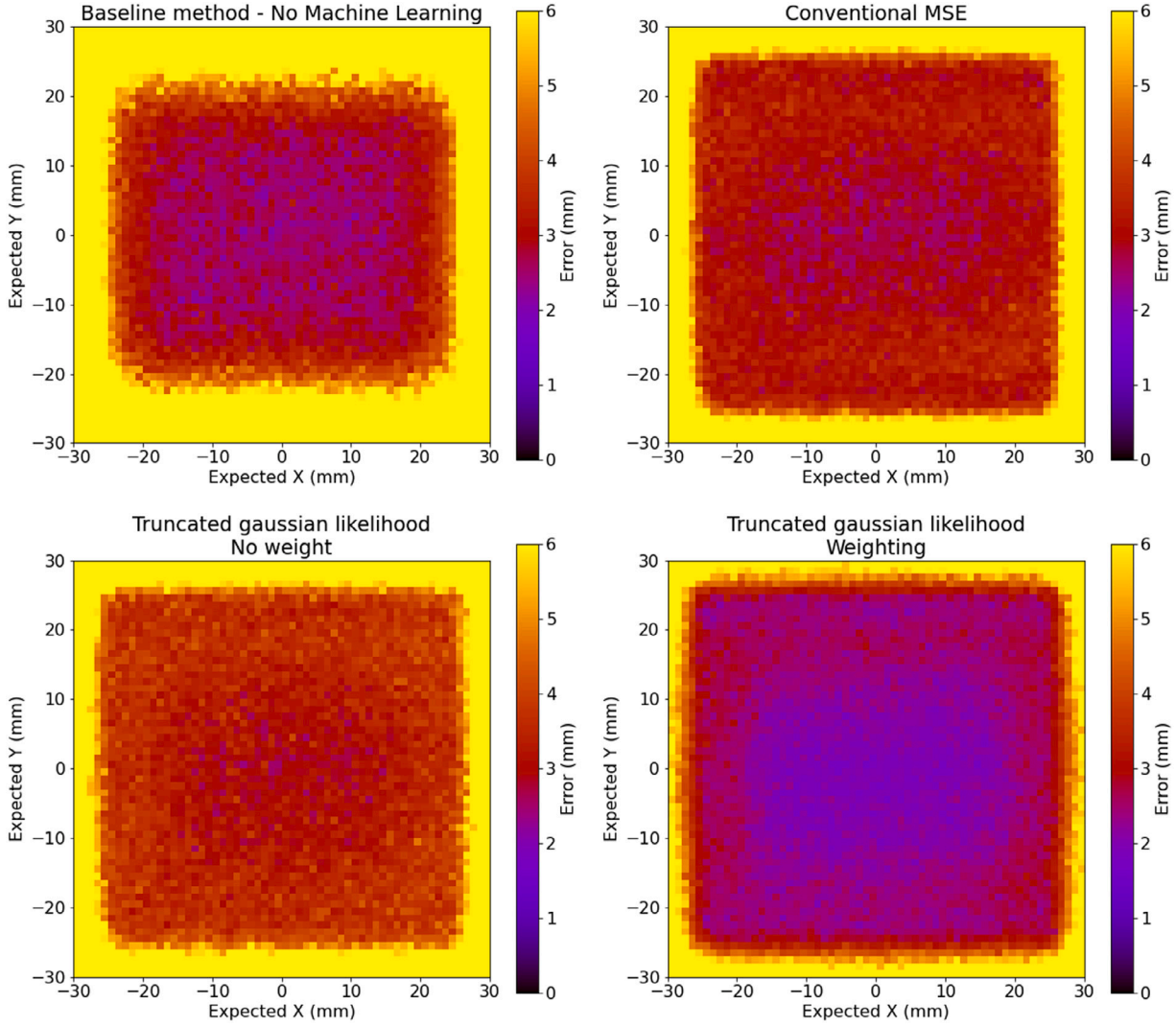

Fig. 7. 2D error reconstruction according to the true position, corresponding to the spread of the reconstruction — The color scale has been threshold at 6 mm for visualization purpose.

performances than the baseline method. The weighting given by our uncertainty estimation outperforms the other methods and reaches a Mean 2D error of 2.51 mm.

4.3. Calibration of the uncertainties

The evaluation of the performances shows the advantage of using the uncertainty prediction as a weighting of the uncertain reconstruction. However, these results do not provide accurate elements to assess the calibration of the uncertainties. In this section, we provide an evaluation of the quality of the uncertainty estimation by using a coverage plot.

To produce this plot, we consider a probability level α. For each event i, our neural network provides a prediction of the parameters $\mu_\theta^{(i)}$ and $\sigma_\theta^{(i)}$ for a truncated Gaussian distribution, with bounds a and b. Under the hypothesis of this distribution, we compute the Prediction Intervals (PI) at level α for each event $I^{(i)}(\alpha)$ such that:

$$p(y_{\text{true}} \in I^{(i)}(\alpha)) = \alpha \tag{13}$$

where y_{true} is the expected value. There is an infinite number of possible intervals that correspond to this condition, we choose the interval that is centered on the expected median. Then, we compute the number of events whose expected value belongs indeed to the Prediction Interval, which leads to the Prediction Interval Coverage Probability (PICP) at level α:

$$\text{PICP}(\alpha) = \frac{|\{y_{\text{true}} \in I^{(i)}(\alpha)\}_i|}{N} \tag{14}$$

where $|E|$ is the cardinal of the ensemble E and N is the total number of events. The PICP can be seen as the empirical frequency of the true values that indeed belong to the Prediction Interval. Finally, we compare the PICP(α) to the probability level α:

- if PICP(α) = α, we have a perfect calibration;
- if PICP(α) > α, the model is under-confident, or in other word, conservative;
- if PICP(α) < α, the model is over-confident.

Fig. 11 shows the results for different probability levels α. For both coordinates, the PICP is close to the value α which means the Prediction Interval can be trusted in average. The model is slightly over-confident for probability levels under 75% and conservative for higher probability levels. For instance, Prediction Intervals at 95% are conservative.

5. Discussion

The use of Machine Learning approaches to reconstruct gamma photon position of interaction in monolithic crystals is a hot topic

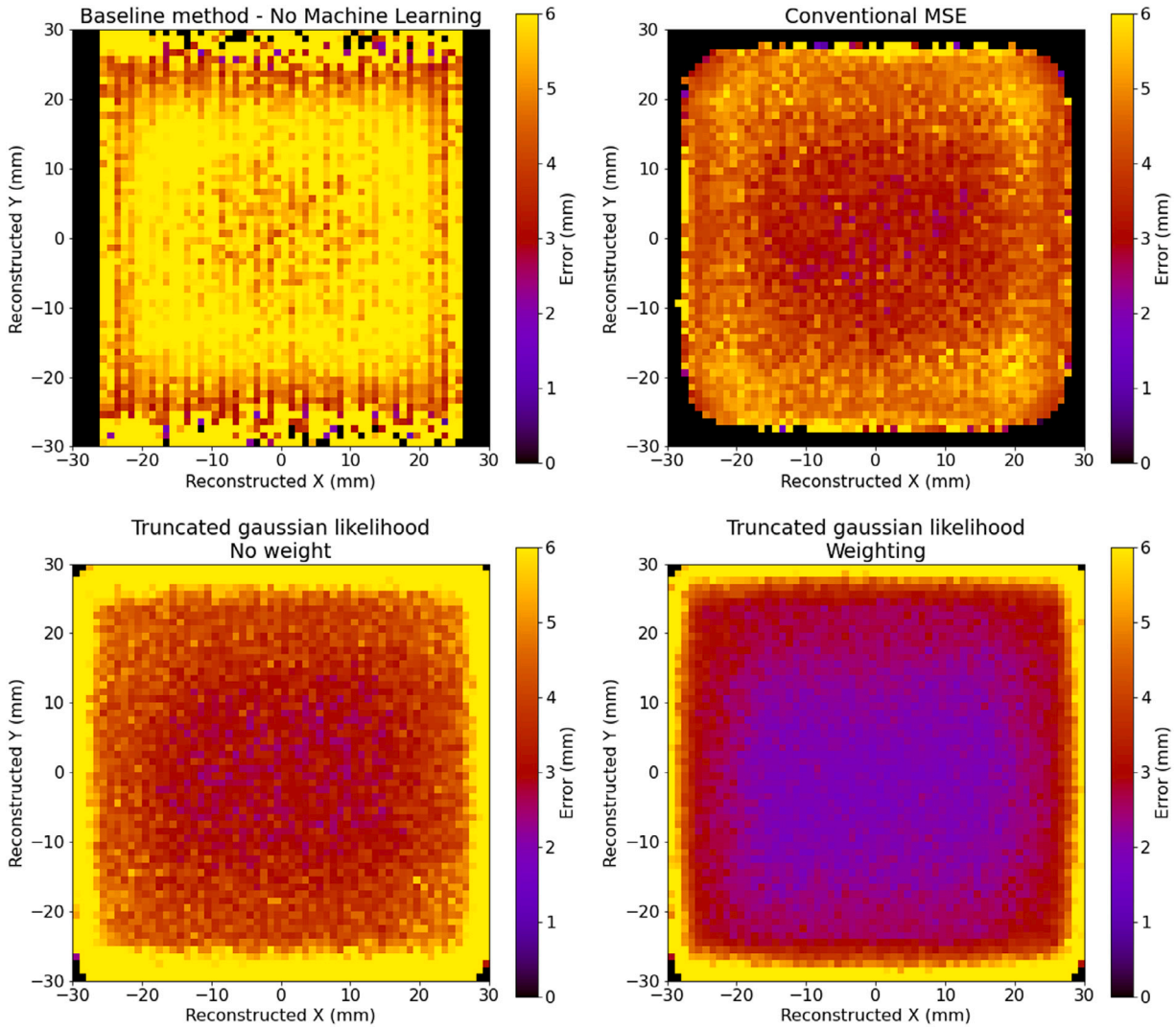

Fig. 8. 2D error reconstruction according to the predicted position, corresponding to the precision of the reconstruction — The color scale has been threshold at 6 mm for visualization purpose.

nowadays. For PET applications, most of the time it is a LYSO scintillator of several tenth of millimeters size and about 1 cm thickness coupled with pixelated silicon photomultiplier (SiPM). For example, in Belov et al. (2023) the spatial resolution obtained with a feed-forward neural network is 0.74 ± 0.01 mm for the XY plane and an average 1.01 ± 0.01 mm for the Z coordinate. Similar results were obtained in Freire et al. (2022) with one multilayer perceptron with five hidden layers and 100 nodes. In Carra et al. (2022), the sub-millimeter precision in XY plane is obtained with an algorithm based on neural networks integrated into a second neural network for simultaneous estimation of the event position and timestamp. Even a preclinical PET system with an annular monolithic LYSO scintillator was proposed by Jaliparthi et al. (2021) with an inner diameter of 6 cm with minimum transaxial scintillator thickness of about 1 cm covered by SiPMs. Authors of these study obtained the spatial resolution on reconstructed interaction position of about half of millimeter with a ten-layer deep residual-convolutional neural networks. Sub-millimeter precise photon interaction position determination in $CeBr_3$ crystal and $LaBr3_3$:Ce (Kawula et al., 2021) for Compton Camera imaging system was obtained with a newly designed convolutional neural network of five layers.

As shown in these recent studies and in Section 4, the Machine Learning approaches provide promising and more accurate reconstructions than the baseline methods, only based on the knowledge of the detector. In contrast to prior research with LYSO or other scintillation only crystals, this current study aims a precise reconstruction of the gamma photon interaction in a $PbWO_4$ crystal with faster (Cherenkov effect) but fewer (lower light yield) optical photons. Moreover, in contrary to pixelated SiPM with one readout channel per pixel, a MCP-PMT with fewer readout channels (2 × 32 instead of 32 × 32) is used. These differences introduce additional complexity on the precise position reconstruction. The coordinate along the transmission lines (X-coordinate) turned out to be complex to reconstruct, but our machine learning models were able to learn to address this complexity.

For monolithic crystals, the so-called edge-effect worsens the spatial localization performance toward the detector borders (Freire et al., 2022). In our study, the use of a dedicated loss function, the truncated Gaussian likelihood, helps recovering some dead area at the edges that are not reconstructed by the baseline methods or the conventional MSE. This effect can be explained by the fact that no interaction can be observed outside of the detector, so the MSE avoids to provide predictions close to the area outside of the detector and consequently, close to the edges. Because we include this prior knowledge in the truncated Gaussian likelihood, the model trained with this loss function can predict positions close to the edges.

Different from previous research, an uncertainty on the gamma photon interaction reconstruction is provided in this study by the Density Neural Network. This additional information is included by a weighting

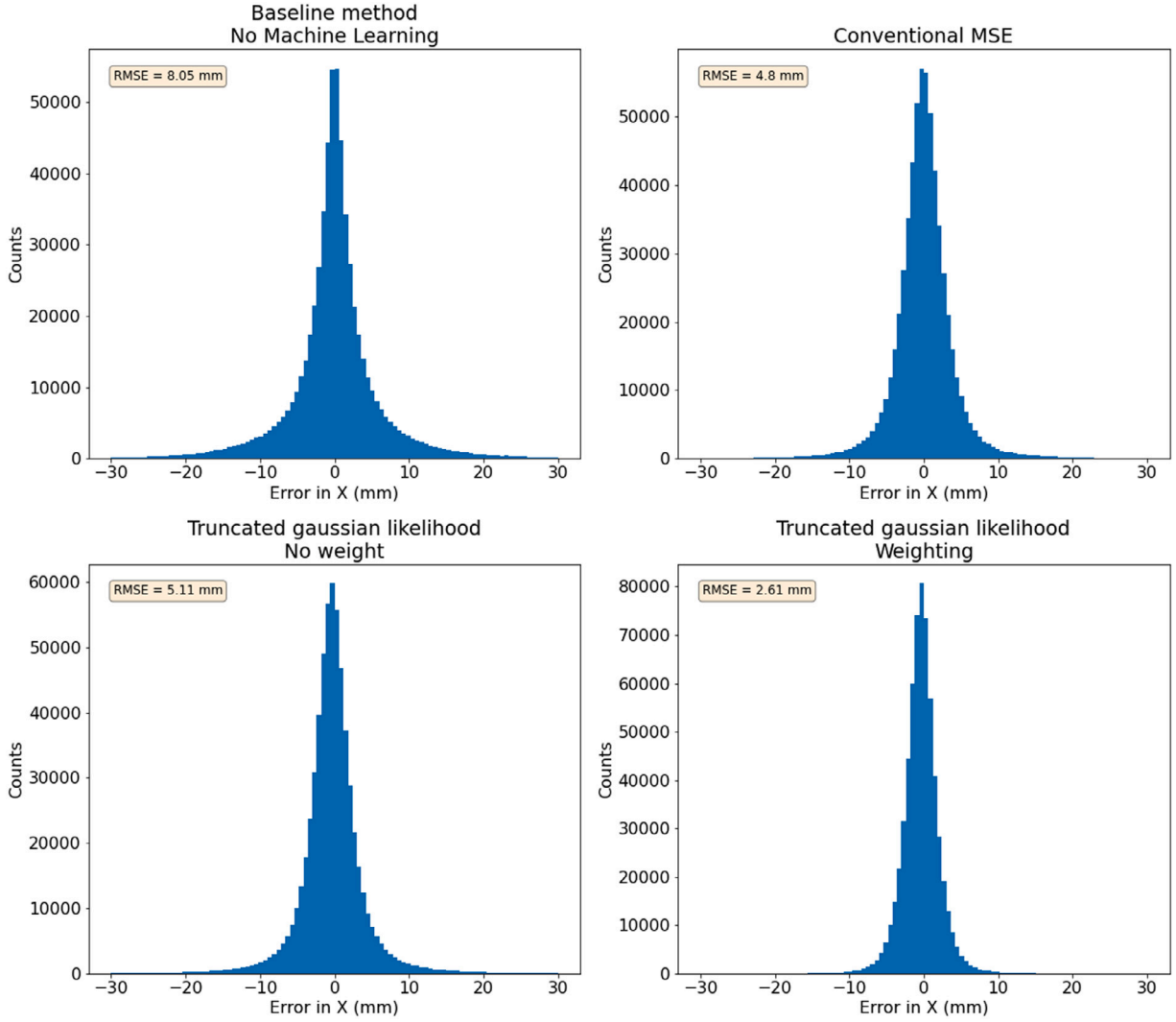

Fig. 9. Global histogram of reconstruction errors on the X position.

in the performance evaluation. We have shown that it significantly improves the performances. We can object that this improvement is "artificial" because we do not improve the prediction, we give less importance in the metrics to the uncertain prediction. However, it is very promising for the application in PET imaging because the tomographic image reconstruction relies on the processing of a large number of events. If some events carry less information, they can degrade the SNR in the reconstructed image. As a consequence, accounting for the uncertainty information in the tomographic reconstruction shall increase the SNR in the image. The uncertainty can be included in a spatial resolution model adapted to each individual event, enabling an event-based resolution modeling image reconstruction. The proposed approach would be innovative for PET imaging, with the best of our knowledge, and its impact on the reconstructed image quality must be assessed. The combination of the proposed gamma photon interaction reconstruction in the detector with the full PET image reconstruction is planned for future works.

Finally, the calibration of the uncertainties show that their prediction is satisfying, although not perfect. The disagreement between the probability level and the empirical frequency can come from two factors:

- the prediction of the neural networks can be improved by optimizing the model and the training phase;
- the truncated Gaussian hypothesis could not be the best representation of the uncertainty. Further works are planned to use other probability distribution functions, under the constraint that these probability distributions must not diverge and be differentiable so that they can be used as loss functions.

6. Conclusion

Our objective to predict the 2D position of gamma interaction in the ClearMind detector is successfully achieved thanks to Machine Learning methods. Our neural network outperforms the baseline approaches based on physical knowledge only and used as inputs of the model. The introduction of Density Neural Networks provides a reliable estimation of the uncertainty in the prediction of the neural network that can be used to discriminate reconstructions that are likely less accurate. We aim to draw benefits from this method in further works for PET image reconstruction. Moreover, we exploit the flexibility of the Density Neural Network approach to design a truncated gaussian loss function based on physical constraints. This property helps for the reconstruction of events in areas far from the center of the detector, increasing the exploitable surface of the crystal. We want to highlight that this methodology is generic and can be adapted to other use cases, as far as we are able to introduce such constraints in the loss function.

The preprocessing step is also important to reduce the number of input variables, thus to get a compact neural network that can be embedded as close as possible to the detector. The computation time on test data on a conventional CPU is 14 s for the 600 000 events, corresponding to 23 μs per event.

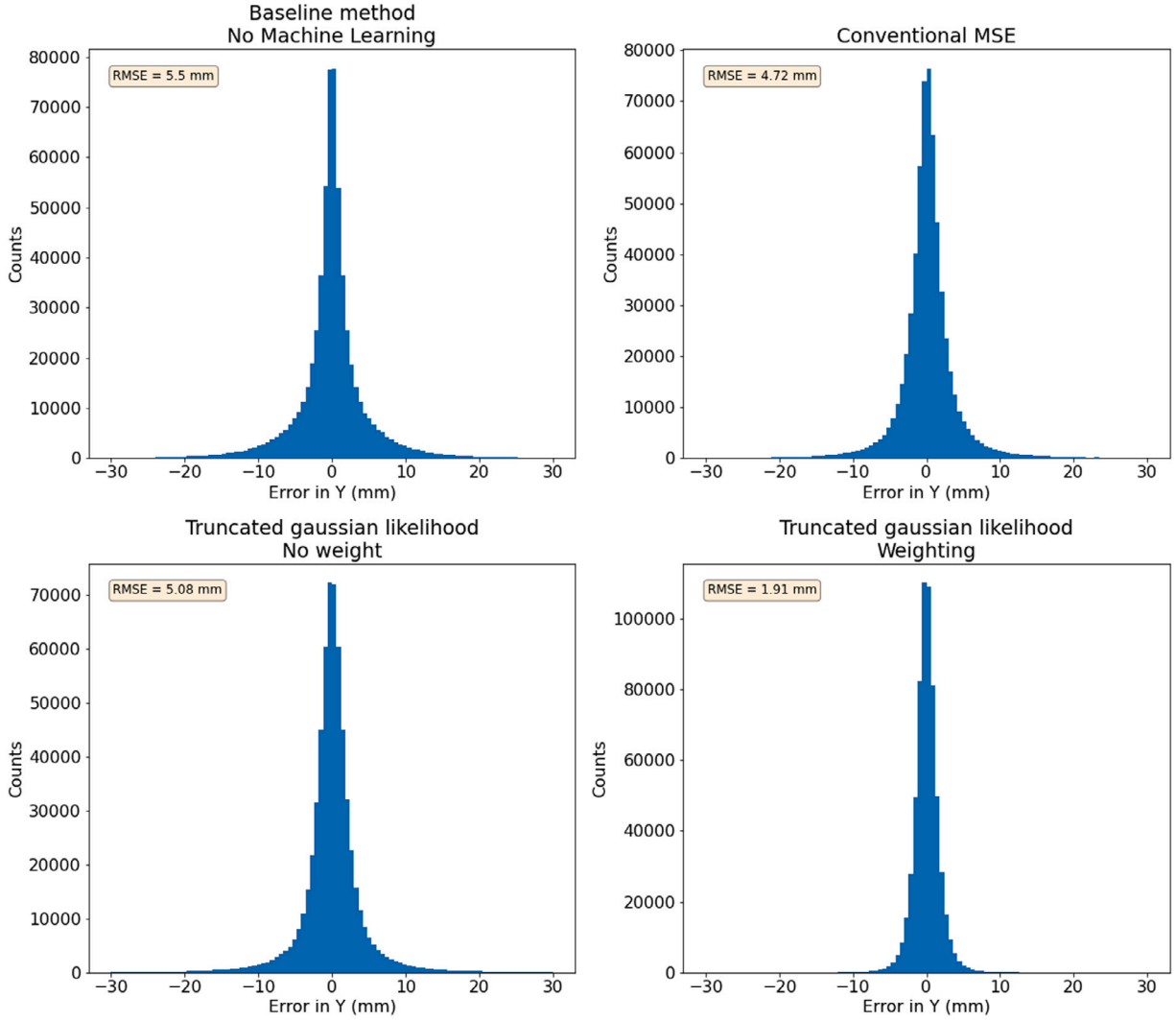

Fig. 10. Global histogram of reconstruction errors on the Y position.

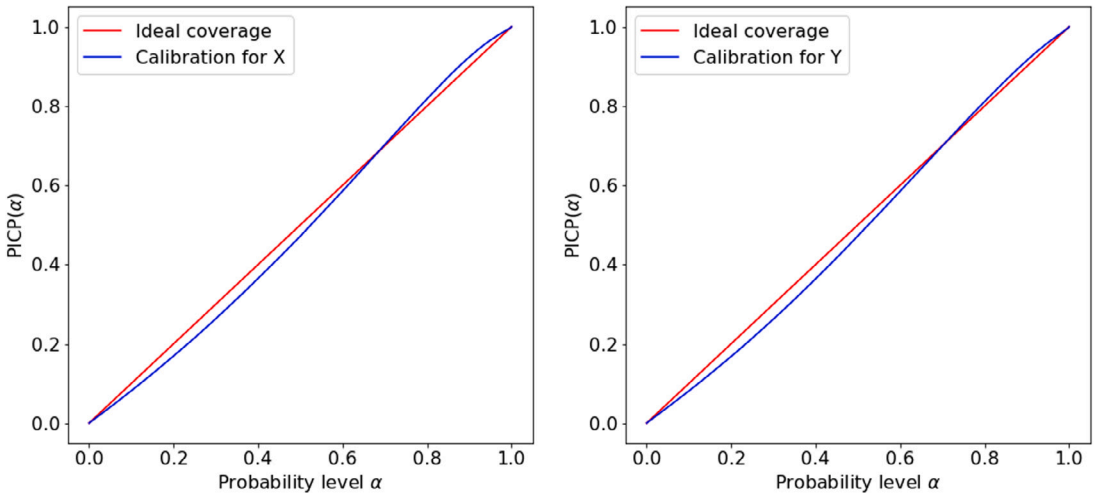

Fig. 11. Coverage plots — The models are slightly too confident for probability levels lower than 75% and slightly too conservative for probability levels higher than 75%.

In outlooks, we want to conduct a sensitivity analysis on the input variables. The objective is to study the importance of the preprocessed observables for the prediction of the positions. We will verify if we find the expected correlation between the prediction of the positions and the identified observables given by the physical expertise. Moreover, the analysis will be performed on the uncertainty estimation in order to understand which quantities have the highest influence on the uncertainty in the reconstruction. We will compare the performances in reconstruction with the use of the raw data, in order to assess the possible loss of information due to this preprocessing. In further works, we will also apply this methodology for the Depth of Interaction, energy and time reconstructions in order to obtain complete information on the gamma ray reconstructions. Moreover, these results will be exploited to complete PET image reconstruction in dedicated studies, in order to improve the SNR on the image by the weighting or the discrimination of uncertain reconstruction.

CRediT authorship contribution statement

G. Daniel: Methodology, Software, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. **M.-B. Yahiaoui:** Methodology, Software. **C. Comtat:** Conceptualization, Writing – original draft, Writing – review & editing. **S. Jan:** Conceptualization, Writing – original draft, Writing – review & editing, Funding acquisition. **O. Kochebina:** Conceptualization, Writing – original draft, Writing – review & editing. **J.-M. Martinez:** Methodology, Writing – original draft. **V. Sergeyeva:** Methodology, Software. **V. Sharyy:** Conceptualization, Software, Data curation, Visualization, Writing – original draft, Writing – review & editing. **C.-H. Sung:** Software. **D. Yvon:** Conceptualization, Software, Data curation, Visualization, Formal analysis, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Sebastien JAN reports financial support was provided by Horizon 2020.

Data availability

Data will be made available on request.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 800 945. We are grateful for the support and seed funding from the CEA, Programme Exploratoire Bottom-Up, under grant No. 17P103-CLEAR-MIND, the Cross-Disciplinary Program on Numerical Simulation of CEA for the financial support of project AAIMME, and the French National Research Agency under grant No. ANR-19-CE19-0009-01.

Appendix A. Statistical processing of raw waveforms

The following paragraphs describes the preprocessed observables we developed for the ClearMind detector in addition to the description in Section 2.3. First, for each line that triggered the SAMPIC acquisition, let us define $F_{l,\text{Left}}(t_j)$ and $F_{l,\text{Right}}(t_j)$, the digitized pulse shapes (Left and Right of transmission line) at time t_j.

Then, before computing event observables, we calculate for each triggered line:

- The times of the first samples of the pulse shapes $T_{0,\text{Left}}$, $T_{0,\text{Right}}$
- The times calculated on the rising edges of the pulses, extrapolated to the half height between the sampled values, $T_{l,\text{Left}}$ and $T_{l,\text{Right}}$.
- The time lapse when the pulses have exceeded the WaveCatcher trigger threshold.
- The time lapse when the pulses have saturated the WaveCatcher.
- The sum of the charges collected at both ends of the transmission lines C_l.

The selected observables can be classified into several categories.

First, a general parameter stores the number of transmission lines that have acquired a signal.

Then we store the observables expected to be correlated to the *interaction time* of the gamma photon in the crystal. These are :

- The time of the first SAMPIC sample recorded on the first transmission line that acquired a signal. This time T_{First} is then subtracted from all the time variables associated with the event.
- The time of the first photoelectron detected on the event. We store the lowest photoelectron time value (defined as $0.5 \times (T_{l,\text{Left}} + T_{l,\text{Right}})$ of all the lines that triggered the acquisition.

We add four bias observables, intended to help the neural network to decorrelate possible stacking effects on the pulse shapes. These are

- the time lapse beyond 50% of the amplitude on the pulse shapes $F_{l,\text{Left}}(t_j)$) and $F_{l,\text{Right}}(t_j)$ (one observable for Left pulse shape and one for Right)
- the computed pulses' rise time (one observable per side)

Then we store the observables expected to be correlated to the *energy deposited* by the interaction in the detector which is a quantity that is planned to be reconstructed for future works. These observables are also expected to bring information to estimate the uncertainties on the interaction position. We store :

- The sum of the charges collected over all lines C_T.
- The sum of the time lapse when pulses have exceeded the triggering threshold of the WaveCatcher T_{Thres} over all lines.
- The sum over of the time lapse when pulses have saturated the WaveCatcher (bias observable), T_{Sat}.

Then we store observables correlated to the position on the axis perpendicular to the transmission lines x_{pos} and along transmission line y_{pos}, as explained in Section 2.3.

Finally, we compute quantities expected to be correlated to the *depth of interaction* (DOI) of the gamma in the crystal, used for future works to achieve 3D reconstruction. We use the fact that the farther the gamma interaction occurs from the photoelectric layer, the more photoelectrons are produced over a large area. The depth of interaction is thus correlated to the dispersion of the produced photoelectrons. We can also expect that this dispersion is relevant for our neural network to estimate the uncertainty on x_{pos} and y_{pos}. To quantify this dispersion we first calculate the distribution: $|l - \text{Med}_L|$, weighted by the charge measured on the line l, C_l. $\text{Med}_l = \text{median}(|l - Med_L|, C_l)$. The two first observables are then :

- the $\text{median}(|l - \text{Med}_L|, C_l)$ and
- the $\text{mean}(|l - \text{Mean}_L|, C_l)$

To quantify the dispersion along the transmission lines, we use the same algorithms, but on the previously computed $\Delta t_{l,p}$ and $\text{Bipol}_{l,p}$ distributions (Section 2.3), weighted by the charge collected in the pulses. The observables are then :

- the $\text{median}(|\Delta t_{l,p} - \text{Med}_{\Delta t}|, C_{l,p})$,
- the $\text{mean}(|\Delta t_{l,p} - \text{Mean}_{\Delta t}|, C_{l,p})$,
- the $\text{median}(|\text{Bipol}_{l,p} - \text{Med}_{\text{Bipol}}|, C_{l,p})$
- the $\text{mean}(|\text{Bipol}_{l,p} - \text{Mean}_{\text{Bipol}}|, C_{l,p})$

We did not devise until now any bias observables on the depth of interaction parameter.

Appendix B. Median algorithm on weighted distributions

Following is our algorithm for computing the median of a binned distribution, written in C++. It uses the STL library to sort the input values. Instead of returning the central value of the bin where the mid-point value of the distribution happen, we decided to make use of the values of the neighboring bins to qualify the result.

```
bool CompPairValues(const PairValues& Pair1, const PairValues& Pair2)
{ return (Pair1.Val<Pair2.Val); }

// Compute the mid point of a set of values VValue of weigths Vweight
double Calc_Median_Weighted( vector<double> Vvalues, vector<double> Vweigth, bool PrintInside) {
 unsigned long NbPair = Vvalues.size();
 if(NbPair==0) { cerr<<"Error in DY_Calc_Median_Weighted: Vvalues empty"<<endl;
  exit(-1);
}

// We store values in pairs
 vector<PairValues> VPair(NbPair);
 double Half_SomWeight = 0.;
 for( unsigned long ind= 0; ind < NbPair; ind++) {
  VPair[ind].Val = Vvalues[ind];
  VPair[ind].weigth = Vweigth[ind];
  Half_SomWeight += VPair[ind].weigth;
 }
 Half_SomWeight *= 0.5;

 // We sort values, stable_sort ou sort
 std::stable_sort( VPair.begin(), VPair.end(), CompPairValues);

// We find mid-weight point
 double ValPlus = 0.;  double ValMinus = 0.;
 double SomWeigPlus = 0.; double SomWeigMinus = 0.;
 unsigned long IndexPlus =0;
 for( unsigned long Index =0; Index<NbPair; Index++)
 { SomWeigPlus+= VPair[Index].weigth;
  ValPlus= VPair[Index].Val;
  IndexPlus= Index;
  if(SomWeigPlus>Half_SomWeight) break;
  SomWeigMinus = SomWeigPlus;
  ValMinus = ValPlus;
 }

 // Compute the Median
 double ValueMedian, ValueMedian2, Median;
 double SomWeigMinDes ;
 if( (IndexPlus == 0) || (IndexPlus == (NbPair-1)) )
  Median = ValPlus;
 else {
  double Un_WeightIndPlus= 1./VPair[IndexPlus].weigth;
  // Rising Option
  double DiffSom = Half_SomWeight-SomWeigMinus;
  double DeltaVal = ValPlus-ValMinus;
  ValueMedian = ValMinus + DeltaVal*DiffSom*Un_WeightIndPlus;
  // Descending option
  double ValMinDes = VPair[IndexPlus+1].Val;
  SomWeigMinDes = Half_SomWeight*2. - SomWeigPlus;
  double DiffSom2 = Half_SomWeight - SomWeigMinDes;
  double DeltaVal2 = ValPlus-ValMinDes;
  ValueMedian2 = ValMinDes + DeltaVal2*DiffSom2*Un_WeightIndPlus;
  Median = 0.5*(ValueMedian+ValueMedian2);
 }
 if( PrintInside) // Print everything for debugging { }
 return Median;
}
```

References

Abadi, Martín, Agarwal, Ashish, Barham, Paul, Brevdo, Eugene, Chen, Zhifeng, Citro, Craig, Corrado, Greg S., Davis, Andy, Dean, Jeffrey, Devin, Matthieu, Ghemawat, Sanjay, Goodfellow, Ian, Harp, Andrew, Irving, Geoffrey, Isard, Michael, Jia, Yangqing, Jozefowicz, Rafal, Kaiser, Lukasz, Kudlur, Manjunath, Levenberg, Josh, Mané, Dandelion, Monga, Rajat, Moore, Sherry, Murray, Derek, Olah, Chris, Schuster, Mike, Shlens, Jonathon, Steiner, Benoit, Sutskever, Ilya, Talwar, Kunal, Tucker, Paul, Vanhoucke, Vincent, Vasudevan, Vijay, Viégas, Fernanda, Vinyals, Oriol, Warden, Pete, Wattenberg, Martin, Wicke, Martin, Yu, Yuan, Zheng, Xiaoqiang, 2015. TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

Agostinelli, S., Allison, J., Amako, K., Apostolakis, J., Araujo, H., Arce, P., Asai, M., Axen, D., Banerjee, S., Barrand, G., Behner, F., Bellagamba, L., Boudreau, J., Broglia, L., Brunengo, A., Burkhardt, H., Chauvie, S., Chuma, J., Chytracek, R., Cooperman, G., Cosmo, G., Degtyarenko, P., Dell'Acqua, A., Depaola, G., Dietrich, D., Enami, R., Feliciello, A., Ferguson, C., Fesefeldt, H., Folger, G., Foppiano, F., Forti, A., Garelli, S., Giani, S., Giannitrapani, R., Gibin, D., Gómez Cadenas, J.J., González, I., Gracia Abril, G., Greeniaus, G., Greiner, W., Grichine, V., Grossheim, A., Guatelli, S., Gumplinger, P., Hamatsu, R., Hashimoto, K., Hasui, H., Heikkinen, A., Howard, A., Ivanchenko, V., Johnson, A., Jones, F.W., Kallenbach, J., Kanaya, N., Kawabata, M., Kawabata, Y., Kawaguti, M., Kelner, S., Kent, P., Kimura, A., Kodama, T., Kokoulin, R., Kossov, M., Kurashige, H., Lamanna, E., Lampén, T., Lara, V., Lefebure, V., Lei, F., Liendl, M., Lockman, W., Longo, F., Magni, S., Maire, M., Medernach, E., Minamimoto, K., Mora de Freitas, P., Morita, Y., Murakami, K., Nagamatu, M., Nartallo, R., Nieminen, P., Nishimura, T., Ohtsubo, K., Okamura, M., O'Neale, S., Oohata, Y., Paech, K., Perl, J., Pfeiffer, A., Pia, M.G., Ranjard, F., Rybin, A., Sadilov, S., Di Salvo, E., Santin, G., Sasaki, T., Savvas, N., Sawada, Y., Scherer, S., Sei, S., Sirotenko, V., Smith, D., Starkov, N., Stoecker, H., Sulkimo, J., Takahata, M., Tanaka, S., Tcherniaev, E., Safai Tehrani, E., Tropeano, M., Truscott, P., Uno, H., Urban, L., Urban, P., Verderi, M., Walkden, A., Wander, W., Weber, H., Wellisch, J.P., Wenaus, T., Williams, D.C., Wright, D., Yamada, T., Yoshida, H., Zschiesche, D., 2003. Geant4 - a simulation toolkit. Nucl. Instrum. Methods Phys. Res. A 506 (3), 250–303.

Allison, J., Amako, K., Apostolakis, J., Araujo, H., Dubois, P. Arce, Asai, M., Barrand, G., Capra, R., Chauvie, S., Chytracek, R., Cirrone, G.A.P., Cooperman, G., Cosmo, G., Cuttone, G., Daquino, G.G., Donszelmann, M., Dressel, M., Folger, G., Foppiano, F., Generowicz, J., Grichine, V., Guatelli, S., Gumplinger, P., Heikkinen, A., Hrivnacova, I., Howard, A., Incerti, S., Ivanchenko, V., Johnson, T., Jones, F., Koi, T., Kokoulin, R., Kossov, M., Kurashige, H., Lara, V., Larsson, S., Lei, F., Link, O., Longo, F., Maire, M., Mantero, A., Mascialino, B., McLaren, I., Lorenzo, P. Mendez, Minamimoto, K., Murakami, K., Nieminen, P., Pandola, L., Parlati, S., Peralta, L., Perl, J., Pfeiffer, A., Pia, M.G., Ribon, A., Rodrigues, P., Russo, G., Sadilov, S., Santin, G., Sasaki, T., Smith, D., Starkov, N., Tanaka, S., Tcherniaev, E., Tome, B., Trindade, A., Truscott, P., Urban, L., Verderi, M., Walkden, A., Wellisch, J.P., Williams, D.C., Wright, D., Yoshida, H., 2006. Geant4 developments and applications. IEEE Trans. Nucl. Sci. 53 (1), 270–278.

Allison, J., Amako, K., Apostolakis, J., Arce, P., Asai, M., Aso, T., Bagli, E., Bagulya, A., Banerjee, S., Barrand, G., Beck, B.R., Bogdanov, A.G., Brandt, D., Brown, J.M.C., Burkhardt, H., Canal, Ph., Cano-Ott, D., Chauvie, S., Cho, K., Cirrone, G.A.P., Cooperman, G., Cortés-Giraldo, M.A., Cosmo, G., Cuttone, G., Depaola, G., Desorgher, L., Dong, X., Dotti, A., Elvira, V.D., Folger, G., Francis, Z., Galoyan, A., Garnier, L., Gayer, M., Genser, K.L., Grichine, V.M., Guatelli, S., Guèye, P., Gumplinger, P., Howard, A.S., Hřivnáčová, I., Hwang, S., Incerti, S., Ivanchenko, A., Ivanchenko, V.N., Jones, F.W., Jun, S.Y., Kaitaniemi, P., Karakatsanis, N., Karamitros, M., Kelsey, M., Kimura, A., Koi, T., Kurashige, H., Lechner, A., Lee, S.B., Longo, F., Maire, M., Mancusi, D., Mantero, A., Mendoza, E., Morgan, B., Murakami, K., Nikitina, T., Pandola, L., Paprocki, P., Perl, J., Petrović, I., Pia, M.G., Pokorski, W., Quesada, J.M., Raine, M., Reis, M.A., Ribon, A., Ristić Fira, A., Romano, F., Russo, G., Santin, G., Sasaki, T., Sawkey, D., Shin, J.I., Strakovsky, I.I., Taborda, A., Tanaka, S., Tomé, B., Toshito, T., Tran, H.N., Truscott, P.R., Urban, L., Uzhinsky, V., Verbeke, J.M., Verderi, M., Wendt, B.L., Wenzel, H., Wright, D.H., Wright, D.M., Yamashita, T., Yarba, J., Yoshida, H., 2016. Recent developments in Geant4. Nucl. Instrum. Methods A 835, 186–225.

Belov, M.V., Kozlov, V.A., Tskhay, V.S., Zavertyaev, M.V., 2023. Resolution estimation in different monolithic PET detectors using neural networks. Phys. Medica 106, 102527.

Bishop, Christopher M., 1994. Mixture Density Networks. Neural Computing Research Group Report NCRG/94/004.

Breton, D., Cheikali, C., Delagnes, E., Maalmi, J., Rusquart, P., Vallerand, P., 2020. Fast electronics for particle time-of-flight measurement, with focus on the SAMPIC ASIC. Nuovo Cimento C 43, 7.

Carra, Pietro, Giuseppina Bisogni, Maria, Ciarrocchi, Esther, Morrocchi, Matteo, Sportelli, Giancarlo, Rosso, Valeria, Belcari, Nicola, 2022. A neural network-based algorithm for simultaneous event positioning and timestamping in monolithic scintillators. Phys. Med. Biol. 67 (13), 135001.

CRYTUR, spol. s r.o., Na Lukách 2283 511 01 Turnov Czech Republic, 2020. https://www.crytur.com.

Follin, M., Chyzh, R., Sung, C.-H., Breton, D., Maalmi, J., Chaminade, T., Delagnes, E., Schäfers, K., Weinheimer, C., Yvon, D., Sharyy, V., 2022. High resolution MCP-PMT readout using transmission lines. Nucl. Instrum. Methods Phys. Res. A 1027, 166092.

Follin, M., Sharyy, V., Bard, J.-P., Korzhik, M., Yvon, D., 2021. Scintillating properties of today available lead tungstate crystals. J. Instrum. 16 (08), P08040.

Freire, Marta, Echegoyen, Sara, Gonzalez-Montoro, Andrea, Sanchez, Filomeno, Gonzalez, Antonio J., 2022. Performance evaluation of side-by-side optically coupled monolithic LYSO crystals. Med. Phys. 49 (8), 5616–5626.

Gawlikowski, Jakob, Tassi, Cedrique Rovile Njieutcheu, Ali, Mohsin, Lee, Jongseok, Humt, Matthias, Feng, Jianxiang, Kruspe, Anna M., Triebel, Rudolph, Jung, Peter, Roscher, Ribana, Shahzad, Muhammad, Yang, Wen, Bamler, Richard, Zhu, Xiao Xiang, 2021. A survey of uncertainty in deep neural networks. CoRR abs/2107.03342.

Gonzalez-Montoro, Andrea, Gonzalez, Antonio J., Pourashraf, Shirin, Miyaoka, Robert S., Bruyndonckx, Peter, Chinn, Garry, Pierce, Larry A., Levin, Craig S., 2021. Evolution of PET detectors and event positioning algorithms using monolithic scintillation crystals. IEEE Trans. Radiat. Plasma Med. Sci. 5 (3), 282–305.

Grant, Alexander M., Deller, Timothy W., Khalighi, Mohammad Mehdi, Maramraju, Sri Harsha, Delso, Gaspar, Levin, Craig S., 2016. Nema nu 2-2012 performance studies for the sipm-based tof-pet component of the ge signa pet/mr system. Med. Phys. 43 (5), 2334–2343.

Jaliparthi, Gangadhar, Martone, Peter F., Stolin, Alexander V., Raylman, Raymond R., 2021. Deep residual-convolutional neural networks for event positioning in a monolithic annular PET scanner. Phys. Med. Biol. 66 (14), 145008.

Jan, S., Santin, G., Strul, D., Staelens, S., Assié, K., Autret, D., Avner, S., Barbier, R., Bardiès, M., Bloomfield, P.M., Brasse, D., Breton, V., Bruyndonckx, P., Buvat, I., Chatziioannou, A.F., Choi, Y., Chung, Y.H., Comtat, C., Donnarieix, D., Ferrer, L., Glick, S.J., Groiselle, C.J., Guez, D., Honore, P.-F., Kerhoas-Cavata, S., Kirov, A.S., Kohli, V., Koole, M., Krieguer, M., van der Laan, D.J., Lamare, F., Largeron, G., Lartizien, C., Lazaro, D., Maas, M.C., Maigne, L., Mayet, F., Melot, F., Merheb, C., Pennacchio, E., Perez, J., Pietrzyk, U., Rannou, F.R., Rey, M., Schaart, D.R., Schmidtlein, C.R., Simon, L., Song, T.Y., Vieira, J.-M., Visvikis, D., Van de Walle, R., Wieërs, E., Morel, C., 2004. GATE: a simulation toolkit for PET and SPECT. Phys. Med. Biol. 49 (19), 4543.

Kawula, M., Binder, T.M., Liprandi, S., Viegas, R., Parodi, K., Thirolf, P.G., 2021. Sub-millimeter precise photon interaction position determination in large monolithic scintillators via convolutional neural network algorithms. Phys. Med. Biol. 66 (13), 135017.

Kendall, Alex, Gal, Yarin, 2017. What uncertainties do we need in Bayesian deep learning for computer vision? CoRR abs/1703.04977.

Lecoq, Paul, Morel, Christian, Prior, John O, Visvikis, Dimitris, Gundacker, Stefan, Auffray, Etiennette, Križan, Peter, Turtos, Rosana Martinez, Thers, Dominique, Charbon, Edoardo, Varela, Joao, de La Taille, Christophe, Rivetti, Angelo, Breton, Dominique, Pratte, Jean-François, Nuyts, Johan, Surti, Suleman, Vandenberghe, Stefaan, Marsden, Paul, Parodi, Katia, Benlloch, Jose Maria, Benoit, Mathieu, 2020. Roadmap toward the 10 ps time-of-flight PET challenge. Phys. Med. Biol. 65 (21), 21RM01.

MAPMT-253 Multi-Anode MCP-P.M.T. Datasheet, from Photek Inc., 2020. https://www.photek.com/pdf/datasheets/detectors/DS034-Auratek-MAPMT253-Detector-Datasheet.pdf.

Sarrut, David, Bała, Mateusz, Bardiès, Manuel, Bert, Julien, Chauvin, Maxime, Chatzipapas, Konstantinos, Dupont, Mathieu, Etxebeste, Ane, Fanchon, Louise M., Jan, Sébastien, Kayal, Gunjan, Kirov, Assen S., Kowalski, Paweł, Krzemien, Wojciech, Labour, Joey, Lenz, Mirjam, Loudos, George, Mehadji, Brahim, Ménard, Laurent, Morel, Christian, Papadimitroulas, Panagiotis, Rafecas, Magdalena, Salvadori, Julien, Seiter, Daniel, Stockhoff, Mariele, Testa, Etienne, Trigila, Carlotta, Pietrzyk, Uwe, Vandenberghe, Stefaan, Verdier, Marc-Antoine, Visvikis, Dimitris, Ziemons, Karl, Zvolský, Milan, Roncali, Emilie, 2021-05. Advanced Monte Carlo simulations of emission tomography imaging systems with GATE. Phys. Med. Biol. 66 (10), 10TR03.

Schaart, Dennis R., 2021. Physics and technology of time-of-flight PET detectors. Phys. Med. Biol. 66 (9), 09TR01.

Sung, Chi-Hsun, 2022. Simulation and Artificial Intelligence for a Gamma Detector for High Resolution PET Imaging (Ph.D. thesis). Université Paris-Saclay.

Sung, C.-H., Cappellugola, L., Follin, M., Curtoni, S., Dupont, M., Morel, C., Galindo-Tellez, A., Chyzh, R., Breton, D., Maalmi, J., Yvon, D., Sharyy, V., 2023. Detailed simulation for the clearmind prototype detection module and event reconstruction using artificial intelligence. Nucl. Instrum. Methods Phys. Res. 1053, 168357, https://www.sciencedirect.com/science/article/pii/S0168900223003479.

van Sluis, Joyce, de Jong, Johan, Schaar, Jenny, Noordzij, Walter, van Snick, Paul, Dierckx, Rudi, Borra, Ronald, Willemsen, Antoon, Boellaard, Ronald, 2019. Performance characteristics of the digital biograph vision PET/CT system. J. Nucl. Med. 60 (7), 1031–1036.

Yvon, D., Sharyy, V., Follin, M., Bard, J.-P., Breton, D., Maalmi, J., Morel, C., Delagnes, E., 2020. Design study of a "scintronic" crystal targeting tens of picoseconds time resolution for gamma ray imaging: the ClearMind detector. J. Instrum. 15 (07), P07029.