Optimized reconstruction of the position of interaction in high-performances γ -cameras

T. Bossis, M.-A. Verdier, L. Pinot, F. Bouvet, T. Beaumont, D. Broggio, O. Caselles, S. Zerdoud, L. Ménard

▶ **To cite this version:**

T. Bossis, M.-A. Verdier, L. Pinot, F. Bouvet, T. Beaumont, et al.. Optimized reconstruction of the position of interaction in high-performances γ -cameras. NDIP20 – Conference on New Developments in Photodetection, Jul 2022, Troyes, France. pp.167907, ⟨10.1016/j.nima.2022.167907⟩. ⟨hal-04446415⟩

HAL Id: hal-04446415

https://hal.science/hal-04446415v1

Submitted on 8 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright - All rights reserved

Optimized Reconstruction of the Position of Interaction in High-Performances γ-Cameras

T. Bossis[a,b], M-A. Verdier[a,b], L. Pinot[a,b], F. Bouvet[a,b], T. Beaumont[c], D. Broggio[c], O. Caselles[d], S. Zerdoud[d], L. Ménard[a,b]

[a] *Université Paris-Saclay, CNRS/IN2P3, IJCLab, ORSAY, France,*
[b] *Université Paris-Cité, IJCLab, ORSAY, France,*
[c] *IRSN, LEDI, FONTENAY-AUX-ROSES, France,*
[d] *IUCTO, ICR, TOULOUSE, France,*

Abstract

Vectorized internal radiotherapy is an efficient modality for cancer treatment but requires a personalized dosimetry to adapt the administered dose for each patient, in order to limit the toxicity to organs-at-risk and maximize therapeutic effects. This can be done by performing quantitative imaging of the target organ with high resolution γ-imaging devices in order to evaluate the uptake and biokinetics of the radiotracer. We developed a high-resolution portable γ-camera with a 10×10 cm^2 field of view dedicated to thyroid imaging during diseases treatments with ^{131}I. In addition to the optimization of the detection elements, the quality of the images also depends on the implementation of efficient methods to reconstruct the interaction position of gamma rays in the monolithic scintillator from the measurement of the scintillation light distribution. We present here the results obtained with two different machine learning methods based on experimental data for reconstruction of the γ-rays interaction position. Those methods reach high spatial performances such as millimeter spatial resolution and submilimeter distortion in

the center of the field of view.

Keywords: Gamma-camera, image reconstruction, machine learning, neural networks, dosimetry

1. Introduction

Targeted radionuclide therapy is one of the most widespread treatment modality for benign and malignant thyroid diseases. In order to maximize the therapeutic effects on the target tissues while minimizing toxicity for organs-at-risk, the absorbed dose should be individually estimated for each patient to account for differences in radiopharmaceutical uptake and biokinetics. Accurate individual estimation of absorbed doses for both treatment planning and post-treatment dosimetry verification requires γ-imaging devices that are optimized for high-energy γ-rays and high photon fluences and that can perform repeated measurements at different times before and after treatment administration. To fit these requirements, we proposed to develop a mobile high-resolution γ-camera specifically designed to dosimetry through planar imaging of the radiotracer biodistribution at patient's bedside during the treatment of thyroid diseases with ^{131}I. A first feasibility prototype with a parallel hole collimator and a 5×5 cm^2 field of view (FOV) corresponding to the detection area showed very promising results [1]. We are now developing a fully-operational clinical device (fig. 1). This clinical prototype has a 10×10 cm^2 FOV suited to the size of enlarged thyroids, a millimetric intrinsic spatial resolution, an energy resolution of 8% FWHM at 364 keV (main gamma emission line of ^{131}I) and a counting capability of 200 kcps with negligible dead-time for early imaging after treatment administration.

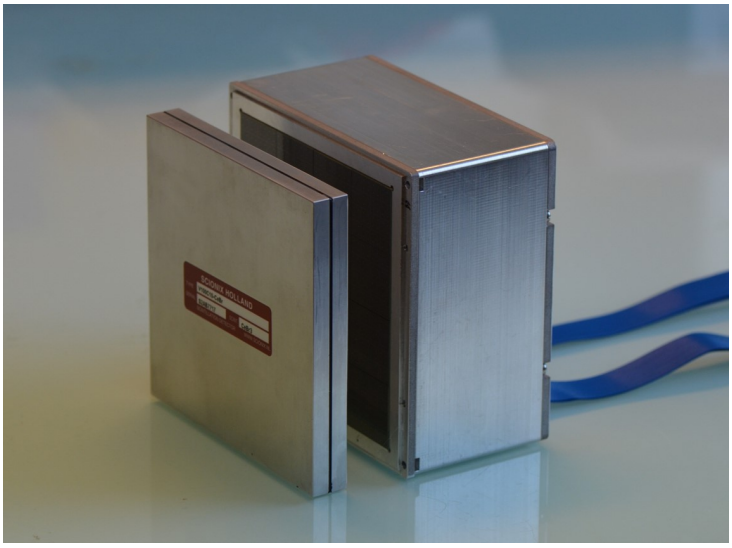

Figure 1: Photodetection module of the mobile camera composed of a 256 SiPMs array coupled to an encapsulated $10 \times 10 \times 1$ cm^3 $CeBr_3$ monolithic scintillator.

Those high intrinsic performances allow to improve the quantitative assessment of radiopharmaceutical distribution far beyond what is possible with conventional γ-cameras, by reducing partial volume effects and contamination from high-energy scattered events. They are firstly achieved through the use of a monolithic inorganic scintillator coupled to pixelated silicon photodetectors, but also requires the implementation of efficient methods to reconstruct the interaction position of γ-rays within the scintillator from the measurement of the scintillation light distribution. Historically, the centroid of the measured light distribution was computed to retrieve the interaction position of absorbed γ-rays [2]. This method is fast but yields poor spatial performance, mainly due to the distortion of the light distribution near the scintillator edges, resulting in a reduced useful field of view (UFOV). It is also possible to define a suitable light distribution analytical model to be fit

on data [3]. Otherwise, if the response function of the detector is unknown or too complex, interpolation methods [4], maximum likelihood [5] or k-Nearest Neighbors [6] based on reference data can be used for reconstruction. These methods however require strong computation, limiting their use for real time clinical applications or when fast processing is required. More recently, the use of Neural networks has shown promising results as they can learn position information from distorted light distributions due to reflections or depth of interaction influence, and therefore, propose an attractive alternative to provide an efficient method for accurate and fast reconstruction of γ-rays. Once trained on a reference dataset, the network can also be used for real-time imaging since the reconstruction is reduced to a few tensor multiplications. This article proposes the optimization and performance comparison of an interpolated model fitting method and a neural network algorithm for γ-ray imaging, both based on experimental data for training and validation.

2. Material and methods

2.1. Global design of the γ-detection system

The γ-detection system of the mobile camera is composed of a 10×10 cm^2 and 1 cm thick $CeBr_3$ monolithic scintillator (Scionix) encapsulated in a mechanical frame with a 1.5 mm thick glass output window and a reflective optical coating. The scintillator is coupled with optical grease (BC-630, Saint Gobain) to a 4×4 Hamamatsu S13361-6050 arrays of 4×4 silicon photomultipliers (SiPMs). The pixels have a sensitive area of 6×6 mm^2 and a microcell pitch of 50 μm. The SiPMs signals are readout by commercial acquisition electronics manufactured by PETSys Electronics. This electronic readout

consists of four TOFPET 2B ASICs with 64 analog readout channels each, used in a charge-integration mode and connected to a DAQ computer via a Gigabit Ethernet link. The SiPMs bias and electronic triggering threshold were adjusted to optimize the energy resolution of the camera. The over-voltage was set to 7.5 V, which corresponds to a gain of $4.3{\cdot}10^6$ and a PDE of 55%. Thermal noise events (11 kcps for the optimal operative bias and individual trigger threshold used) are completely suppressed by an internal hardware trigger (10 ns coincidence window), which operates between two regions of the detector defined as a checkerboard, where each square corresponds to a 4×4 SiPM array. The non-uniformity of the photodetector light response was evaluated and corrected by irradiating the SiPMs with a pulsed LED source. The relative standard deviations of the light response over the 256 SiPMs before and after uniformity correction are 12.8% and 0.96%, respectively.

2.2. Experimental Data

The training and validation of the reconstruction algorithms was conducted with experimental data measured using a ^{133}Ba source collimated with a single-hole tungsten collimator of 0.5 mm diameter . A scan of the whole FOV of the monolithic scintillator-photodetector assembly was performed with a 1 mm step, ranging from -49.5 mm to 49.5 mm in the x and y directions from the center of the camera, and resulting in a 100×100 positions grid. Each interaction event produces a light pulse within the monolithic scintillator that propagates, possibly after some reflections, to the SiPMs array leading to a 16×16 pixels frame corresponding to the charge distribution. The cleaning of the reference dataset is a crucial step before training the

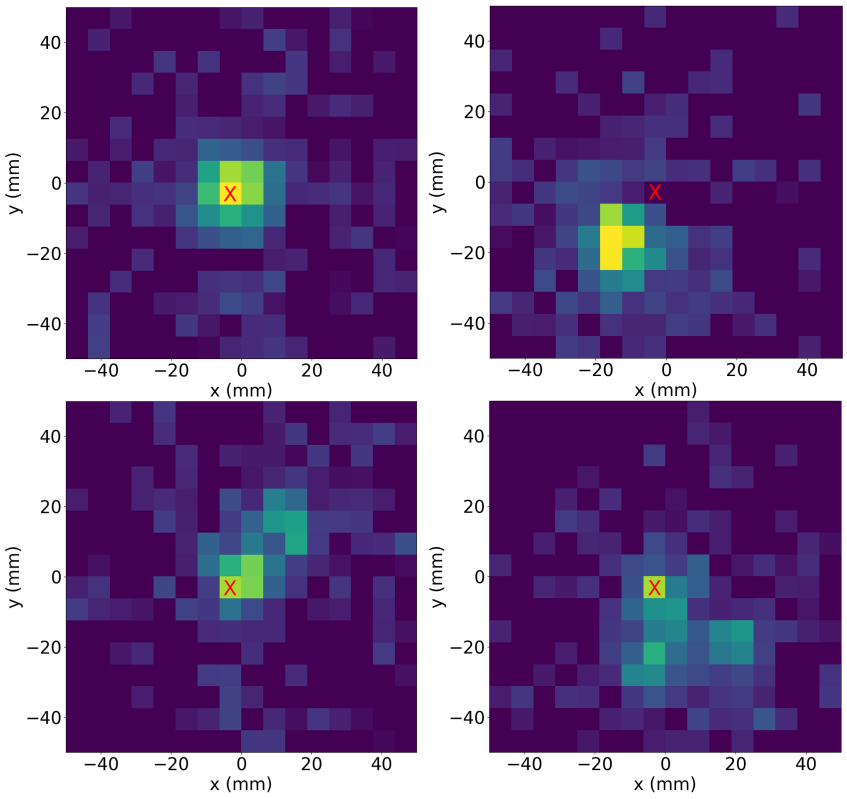

Figure 2: Typical charge distributions produced after the total energy absorption of a 364 keV γ-ray in the monolithic scintillator : (Top-left) photoelectric effect after passing through the hole of the collimator, (Top-right) photoelectric effect after penetrating through the collimator material and (Bottom-left) single Compton or (Bottom-right) multiple Compton scatterings followed by a photoelectric interaction

. The red cross represents the mechanical position of the source.

reconstruction algorithms if good spatial performance is to be achieved. A first energy selection of twice the energy resolution was applied around the 356 keV emission of ^{133}Ba (16% energy window) to only keep events corresponding to a full deposition of the incident energy photon. Fig. 2 shows typical patterns of the frames produced by the full energy absorption of a 356 keV γ-ray in a 1 cm thick $CeBr_3$ scintillator. The scan dataset was then filtered by a second spatial selection to remove full-energy events that have penetrated through the collimator material (fig. 2 top-right) or that have scattered once or several times by Compton interaction inside the scintillator, before being absorbed by photoelectric interaction, which will be called CS+PE events in the rest of the manuscript for concision (fig. 2 bottom left and right). The penetration events represent 77% of all events detected in the energy window, while CS+PE events remain negligible with less than 1% of all detected events. The 22% remaining events correspond to 356 keV gamma rays that propagated through the collimator hole and were fully absorbed inside the scintillator after a single photoelectric interaction (fig. 2, top-left). Both types of spurious events do not produce a well-localized energy deposition and therefore result in poor localization. In order to clean data from parasite events, the position of all events was reconstructed with a non-linear weighting method [2] and only those located within a radius of 3 mm around the mean reconstructed position were kept (the value of the radius was empirically optimized). Two thousand full-energy interactions were then obtained for each scan position after spatial filtering of spurious events. Half of these events were used as a reference dataset for training of the reconstruction methods, while the other half were used as a test dataset

for performance evaluation. Additionally, a flood-field uniformity acquisition has been carried out with the collimatorless ^{133}Ba source, placed at 50 cm from the scintillator (30 million events in the 364 keV photopeak window). The impact on the quality of the reconstruction of CS+PE γ-rays has been quantified. Conversely, their effect on the learning process, that is their inclusion in the training dataset, has not been studied because their number is too small to have any influence.

2.3. Reconstruction algorithms

Two reconstruction methods of the interaction position of the incident γ-ray limited in the $x0y$ plane (no depth of interaction) were tested: a least-square fitting method and a neural network.

2.3.1. Interpolated model fitting

As the response function of each pixel of the camera is unknown, preventing the use of an analytical model for fitting light distributions, the discrete response functions of the pixels were determined at each position of the scan by averaging the frames of the reference dataset. New synthetic light distributions at any given x and y positions of the field of view are interpolated from these reference frames with bivariate spline functions. The interaction position of any detected event of the test dataset is reconstructed by fitting its charge distribution with a standard least-square method (Levenberg-Marquardt), minimizing an objective function calculated at each step thanks to the synthetic light distributions [4]. To reduce statistical fluctuations that may degrade the quality of the fit, the frames of the reference and test dataset can be smoothed with different sizes of gaussian convolution kernel

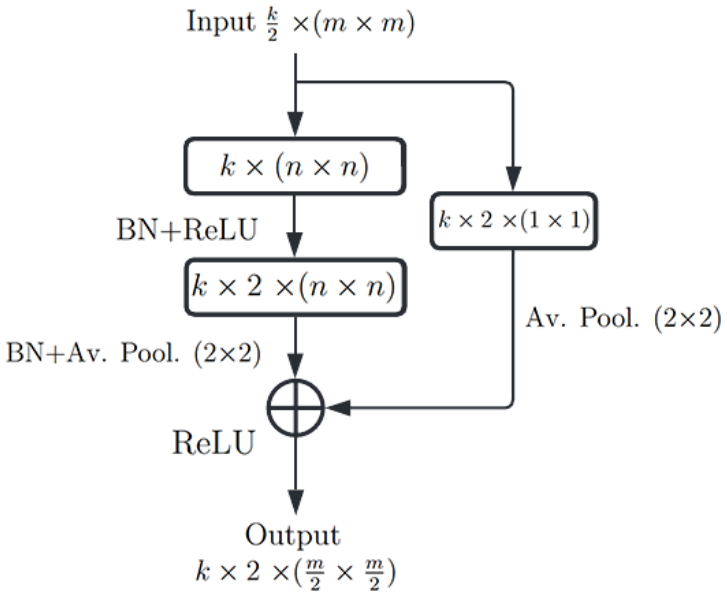

Figure 3: Structure of a *Deep Residual Convolutional Block*. m is the dimension of the input image, n is the dimension of the filters and k is the number of filters. BN stands for Batch Normalization, *Av.Pool.* for Average Pooling and *ReLU* for Rectified Linear Unit.

(3×3, 5×5, ...). The trade-off between precision and convergence speed of the fitter can also be optimized by changing its tolerance for termination parameter [7].

2.3.2. Neural network

The neural network has been designed and trained using Keras (Tensorflow) Python library. It is based on a *Deep Residual Convolutional* architecture inspired from Jaliparthi *et al.* [8]. This architecture is composed of one or several *Deep Residual Convolutional Blocks* each consisting of an input image, directed on one side through two convolutional and Batch Normalization layers, and on the other side to an identity shortcut (skip connection). The outputs of these two branches are then summed and activated by a ReLU (Rectified Linear Unit) activation function, as shown on fig. 3. The use of Batch Normalization is advised to make the training of the neural network faster and more stable [9]. However, when used in a deep neural network, these layers can also cause gradient vanishing or exploding, leading to difficulties of training. This problem is addressed by the use of the skip connections [10]. Unlike Jaliparthi *et al.*, the Leaky ReLU activation function was not used in order not to increase the number of hyperparameters too much and to stay as close as possible to *Deep Residual Convolutional Blocks* defined by He *et al.* [10]. In addition, our images are zero-padded to keep their size $m \times m$ constant during a convolution operation, then passed through an average pooling layer at the end of a block to reduce their size to $\frac{m}{2} \times \frac{m}{2}$. This allows to use more convolution layers by choosing when the image size shrinks. The number of filters k on a layer, that is the number of extracted features, is always two times more than the previous layer, all along

the network, as shown on fig. 3. Therefore, the number of filters on each layer can be derived from the number of filters on the first layer. The size $n \times n$ of the filters is constant in a block but can change from one block to another. The network loss optimization, where the loss can be different functions for regression tasks such as Mean Absolute Error (MAE), Mean Squared Error (MSE) or Mean Squared Logarithmic Error (MSLE), is achieved thanks to the stochastic gradient descent *Adam* algorithm [11]. The neural network is trained thanks to the aforementioned reference dataset, using the 16×16 pixels frames and true mechanical position of the collimated source as input images and labels, and reconstructed interaction positions of the γ-rays as the output, on two fully connected neurons.

2.4. Performance metrics

For both reconstruction methods, the intrinsic spatial performance of the monolithic scintillator-photodetector assembly was evaluated using the second part of the scan dataset, that was not used for training. Those measurements on experimental data do not allow to assess the intrinsic performance of the reconstruction methods because of many sources of physical uncertainty. However, it permits to get relevant information about the interest of those methods for real-world clinical applications. The performance metrics are defined by the NEMA standard [12] and adapted to small field of view γ-cameras by Bhatia *et al.* [13]. The evaluation includes the measurement of the spatial resolution (SR) in mm FWHM by fitting a 2D Gaussian on the 2D histogram of the reconstructed positions, at each scan position. The measured FWHM is deconvoluted from the size of the collimator hole. This fit also gives a measurement of the bias, also called Local Intrinsic Spatial

Distortion (LISD), that is the mean positioning error at each scan position. On the other hand, the integral and differential uniformities (respectively IU and DU) of the spatial response were obtained by reconstructing the data of the flood-field uniformity acquisition. The integral uniformity is given by the standard deviation of counts per bin over the mean counts per bin in the image. It measures how well the counts are uniformly distributed over the whole field of view. The differential uniformity measures local relative variations of counts between adjacent bins for all the bins of the image and the value reported is the mean of all those variations. This metric reflects local accumulations or diminution of counts due to local distortion. All performance are reported in the Central Field of View of the camera (CFOV, 75% of the linear dimension of the FOV), and in the remaining surface uncovered by the CFOV ($\overline{\text{CFOV}}$).

The reconstruction of CS+PE γ-rays can't alter the FWHM spatial resolution as their number is very low in a 1 cm thick scintillator. The effect can only be highlighted by measuring the prediction error, which is the Euclidean distance between the reconstructed position and the expected position. The interquartile range of the prediction error distribution, which is a robust measurement of the dispersion relatively insensitive to outliers, is reported to evaluate the influence of CS+PE events.

2.5. Optimization of the algorithms

The different tested configurations for the fitting method and the neural network are summarized respectively in table 1 and 2. Each time a parameter is changing, all other parameters remain constant and identical to those of the reference configuration *B1* for the fitting method, and *B2* for the neural

Configuration name	Number of events/pos.	Gaussian kernel size	Tolerance for fit termination	Reference scan step
A1	150	3×3	10^{-1}	1 mm
B1	**300**	**3×3**	**10^{-1}**	**1 mm**
C1	600	3×3	10^{-1}	1 mm
D1	900	3×3	10^{-1}	1 mm
E1	300	no smoothing	10^{-1}	1 mm
F1	300	5×5	10^{-1}	1 mm
G1	300	3×3	10^{-2}	1 mm
H1	300	3×3	10^{-4}	1 mm
I1	300	3×3	10^{-1}	3 mm
J1	600	3×3	10^{-2}	3 mm

Table 1: Summary of the different tested configurations for the fitting algorithm. The tolerance for fit termination refers to xtol and ftol parameters of Python's Scipy library documentation [7]. The same value is applied to both parameters. The reference configuration is in bold and the configuration that gives the best compromise in terms of spatial performance is in red.

network. The optimized parameters for the fitting algorithm are the number of events/position of the reference scan, the step of the reference scan, the size of the gaussian smoothing kernel and the convergence tolerance of the fitter [7]. The optimized parameters for the neural network are the number of events/position and the step of the scan for the training dataset, the number of filters per layer, the number of residual blocks, the loss function and the effect of the residual connection. The model has been trained for 30 epochs and saved at each epoch to compare performance. We noticed that for all

Configuration name	Nbr of events/pos.	Nbr of filters on 1^{st} layer	Filter size	Nbr of res. blocks	Reference scan step	Loss function	Residual connection
A2	200	16	3×3/5×5	2	1 mm	MSE	yes
B2	**500**	**16**	**3×3/5×5**	**2**	**1 mm**	**MSE**	**yes**
C2	800	16	3×3/5×5	2	1 mm	MSE	yes
D2	500	8	3×3/5×5	2	1 mm	MSE	yes
E2	500	32	3×3/5×5	2	1 mm	MSE	yes
F2	500	16	3×3/3×3	2	1 mm	MSE	yes
G2	500	16	5×5/5×5	2	1 mm	MSE	yes
H2	500	8	3×3/3×3/5×5	3	1 mm	MSE	yes
I2	500	16	3×3/5×5	2	3 mm	MSE	yes
J2	500	16	3×3/5×5	2	1 mm	MAE	yes
K2	500	16	3×3/5×5	2	1 mm	MSE	no
L2	800	8	5×5/5×5/5×5	3	1 mm	MAE	yes

Table 2: Summary of the different tested configurations for the neural network. MSE stands for Mean Squared Error, and MAE for Mean Absolute Error. The reference configuration is in bold and the configuration that gives the best compromise in terms of spatial performance is in red.

network architectures the loss value reached a plateau around the 17^{th} epoch. We therefore established a learning rate schedule decreasing the learning rate from 10^{-3} to 10^{-4} at the 17^{th} epoch. The batch size is set to 32 samples. As *Adam* optimizer uses a stochastic gradient descent and the weights and bias of the network are randomly initialized, each neural network architecture is trained three times to ensure that the training is stable, and the reported

performances are the mean of the three trained models.

3. Results

The box plots in fig. 4, 5, 7 and 8 represent the distributions of spatial resolution and distortion in the CFOV and $\overline{\text{CFOV}}$. The bottom and the top of the box are respectively the first and third quartiles $Q1$ and $Q3$, so the box height is the interquatile range IQR= $Q3 - Q1$. The bottom and top whiskers represent $Q1 - 1.5{\cdot}\text{IQR}$ and $Q3 + 1.5{\cdot}\text{IQR}$, respectively. The lines crossing the boxes are the median values, the triangles are the mean values, and the dashed line is the lowest median of all configurations. This representation gives an idea of the homogeneity of the spatial response and the possible skewness towards high values. For the uniformity bar graphs on fig. 6 and 9, the blue bar is the uniformity value in the CFOV, the orange bar is the uniformity in the $\overline{\text{CFOV}}$ and the dashed line is the lowest uniformity in CFOV of all configurations.

3.1. Fitting algorithm

The parameters of the fitting algorithm that give the best compromise in terms of spatial performance and processing speed (configuration *J1*) are 600 events per averaged reference frame, a smoothing of those frames and the ones to be reconstructed with a 3×3 gaussian kernel, tolerance for termination of the least-squares fitting of 10^{-2} on the change of the cost function and the change of the independent variables, and a 3 mm step reference scan. These parameters are the ones that, separately, optimize spatial performance. The addition of CS+PE γ-rays to the reconstruction increases the interquartile range of the prediction error distribution by 9%. An example of a 2 mm step

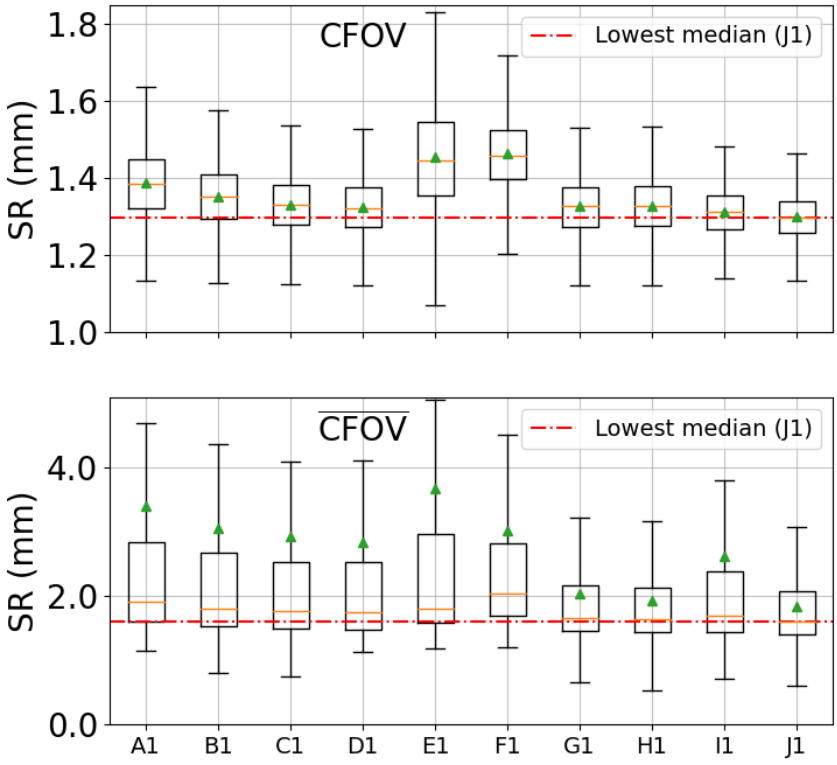

Figure 4: Box plots of the spatial resolution distributions (in mm FWHM) for the different configurations of the fit method. Top: Spatial resolution in CFOV. Bottom: Spatial resolution in $\overline{\text{CFOV}}$.

reconstructed scan with this configuration is shown on fig. 10 (left). The processing speed achieved with this configuration is about 90 events/s/CPU. The overall spatial performance given by this set of parameters is reported in table 3.

3.2. Neural Network

The configuration of the neural network that leads to the best spatial performance uses the optimal value of each separately optimized parameter

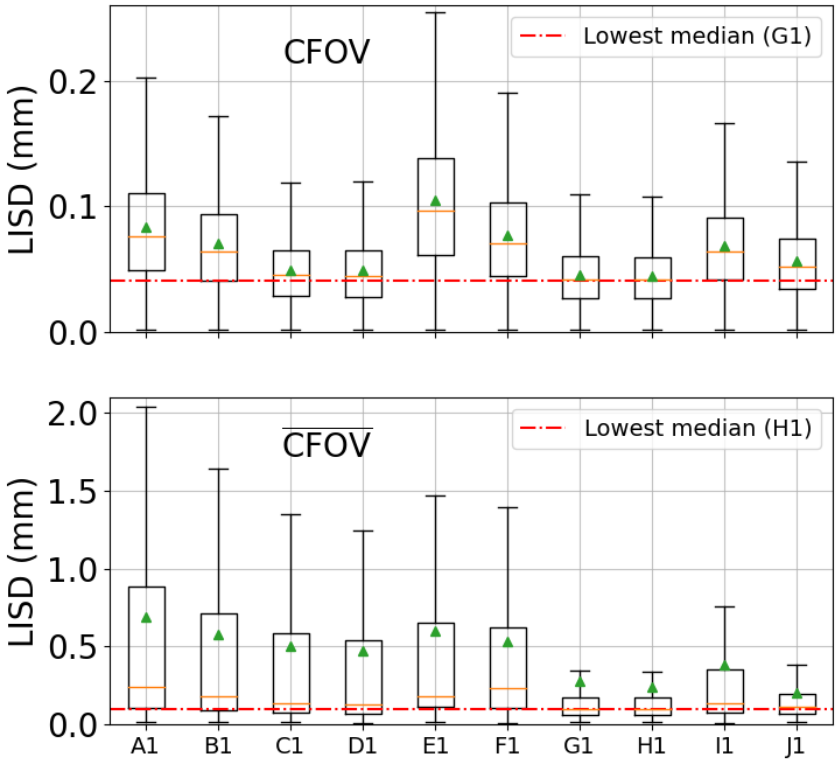

Figure 5: Box plots of the LISD value distributions (in mm) for the different configurations of the fit method. Top: Distortion in CFOV. Bottom: Distortion in $\overline{\text{CFOV}}$.

(configuration *L2*). The training dataset consists of 800 events per position with a step size of 1 mm. The neural network is composed of three *Deep Residual Convolutional Blocks*, where the first block contains 8 and 16 filters, respectively. The second block is similar but with 32 and 64 filters, and the last block uses 128 and 256 filters. All the filters are 5×5 kernels and the network contains a total of around 1 million neurons. We found that the *Mean Absolute Error* loss function leads to better overall performance. The optimal configuration uses residual connections. The influence of CS+PE events is an increase of the interquartile range of prediction error distribution by 9%,

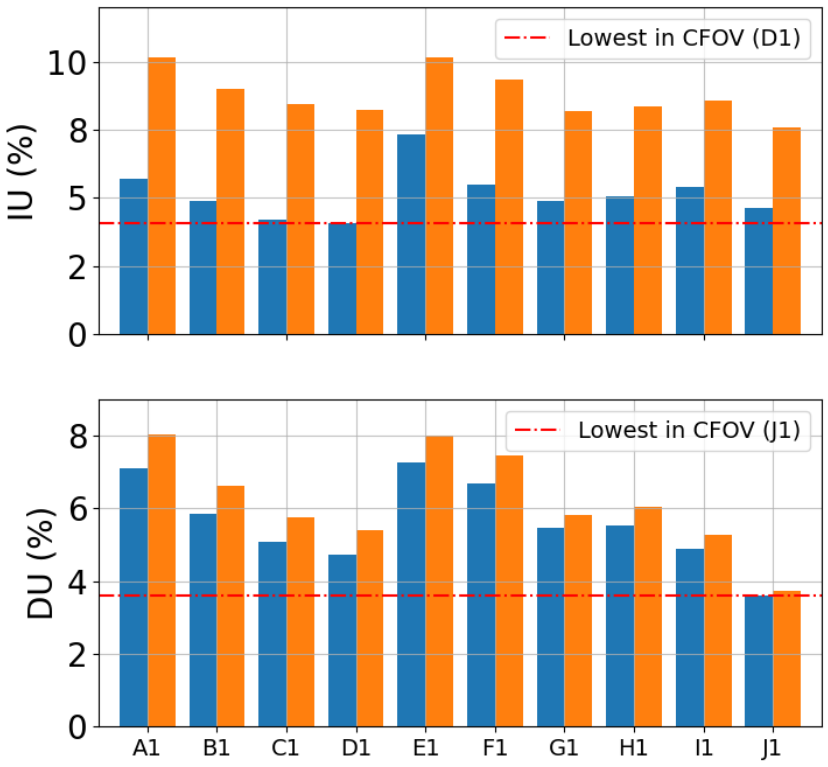

Figure 6: Bar graphs of differential and integral uniformity (in %) for the different parameters configurations of the fitting algorithm. Blue: Uniformity in CFOV. Orange: Uniformity in the $\overline{\text{CFOV}}$.

as for the fitting method. An example of a 2 mm step reconstructed scan with this configuration is shown on fig. 10 (right). The processing speed achieved with this configuration is about 350 events/s/CPU. The overall spatial performance achieved with this architecture is reported in table 3.

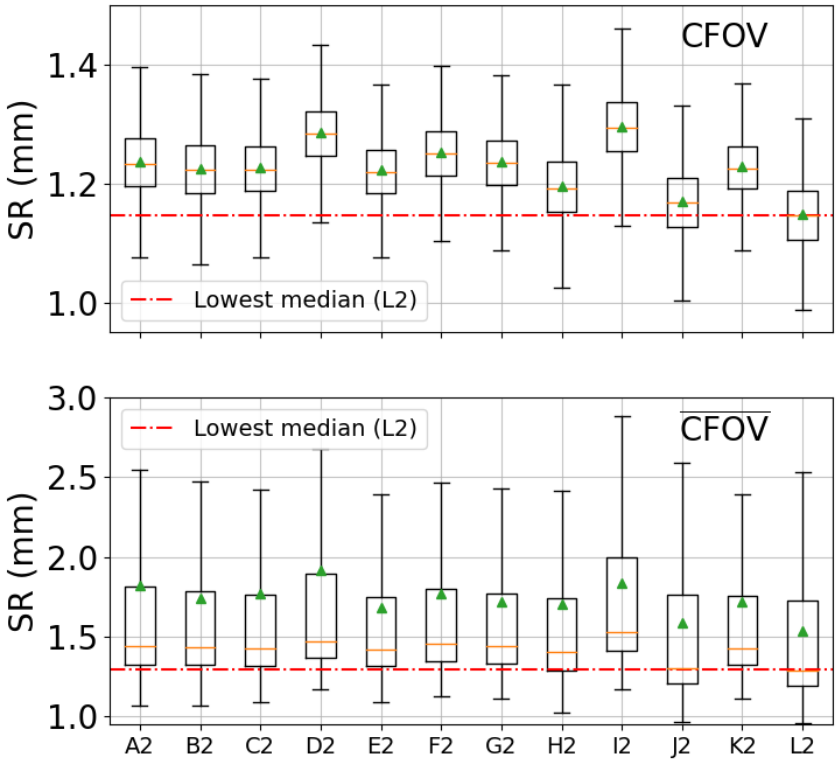

Figure 7: Box plots of the spatial resolution distributions (in mm FWHM) for the different hyperparameters configurations of the neural network. Top: Spatial resolution in CFOV. Bottom: Spatial resolution in $\overline{\text{CFOV}}$.

4. Discussion

4.1. Fitting algorithm

The asymptotic behaviour of the evolution of spatial performance as a function of the number of frames in the computation of the average light responses of SiPMs (*A1* to *D1*) shows that fluctuations become negligible above 600 frames per position and no longer impact the spatial response. Even when reducing the fluctuation on the average frame, it is still necessary to smooth those reference frames along with the frame to be reconstructed

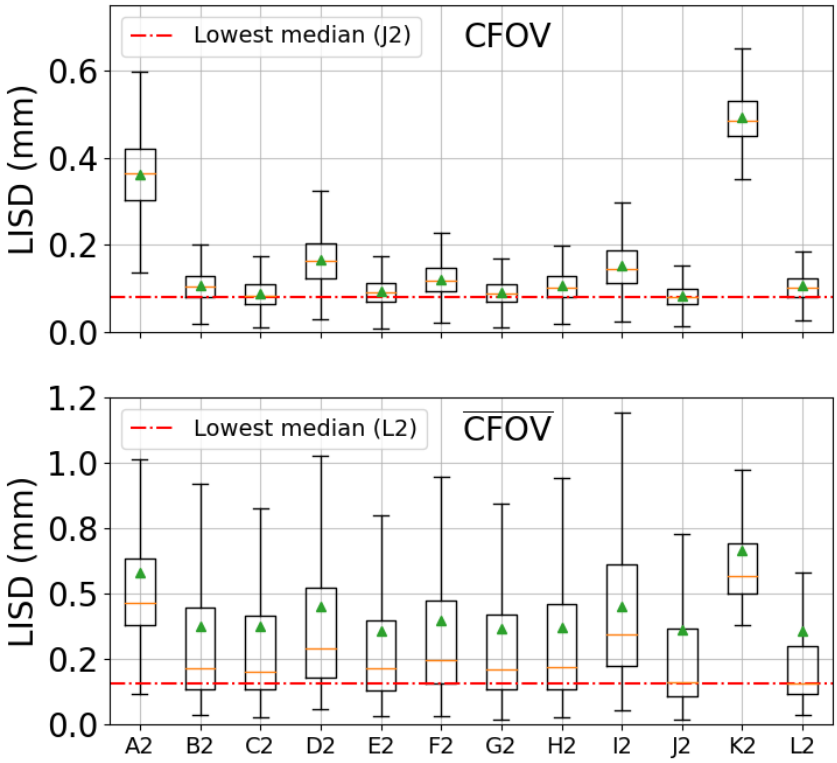

Figure 8: Box plots of the LISD value distributions (in mm) for the different hyperparameters configurations of the neural network. Top: Distortion in CFOV. Bottom: Distortion in $\overline{\text{CFOV}}$.

(*E1*, *B1* and *F1*). The optimal smoothing in terms of spatial performance in CFOV is achieved with a 3×3 gaussian kernel (*B1*). This result may be due to the large size of the SiPMs relative to the scintillator thickness, which results in spatial subsampling of the light distribution and strong pixel-to-neighbor variations. Smoothing reduces the latter effect and also the fluctuations in the tails of the light distributions. However, a too large smoothing (5×5 gaussian kernel, *F1*) decreases the quality of the spatial response. As the surface of the filter represents about 10% of the frame, it mixes useful sig-

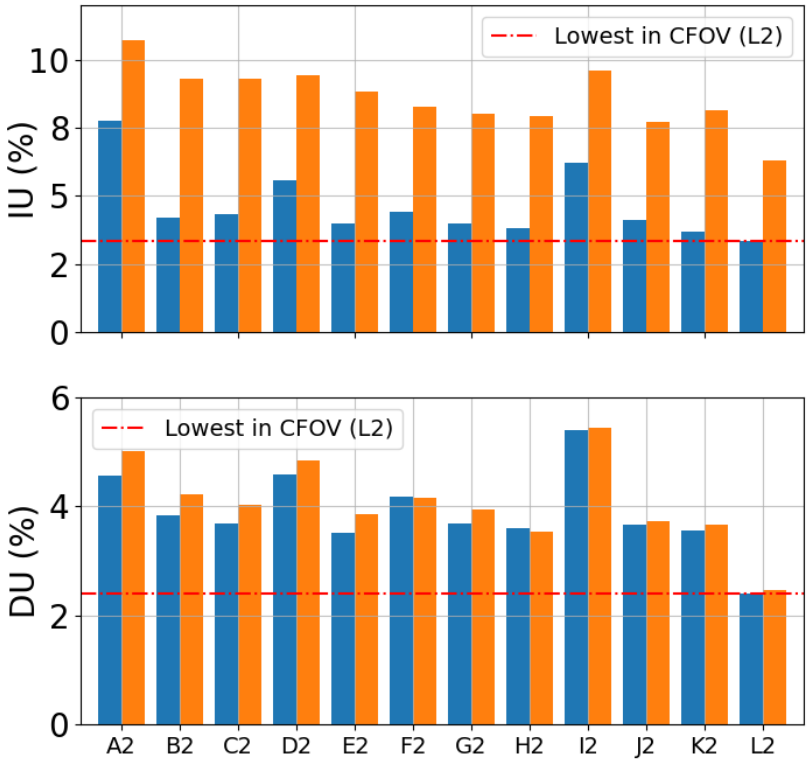

Figure 9: Bar graphs of differential and integral uniformity (in %) for the different hyperparameters configurations of the neural network. Blue: Uniformity in CFOV. Orange: Uniformity in the $\overline{\text{CFOV}}$.

nal information of the light distribution with noise from the tails and thus, degrades the information. Increasing the constraint on the tolerance for termination of the least-squares fit should intuitively improves the accuracy of the reconstruction, which is confirmed by the data. Decreasing the tolerance for termination from 10^{-1} to 10^{-2} (*B1* and *G1*) leads to a significant improvement in spatial resolution and distortion at the edges of the FOV. Further decreasing of this parameter to 10^{-4} (*H1*) is not beneficial and reduces the processing speed of the method of 30% from configuration *G1*. Using a step

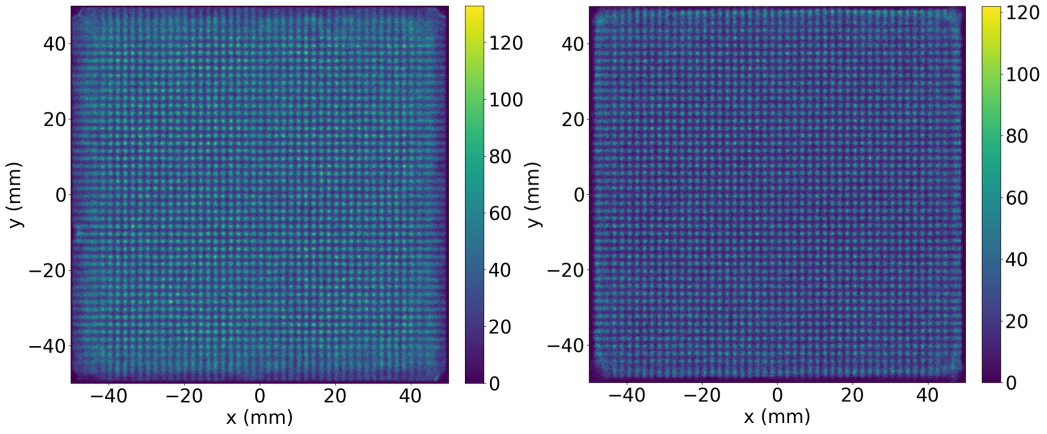

Figure 10: Images of a reconstructed scan obtained with the optimized configurations of the fitting algorithm (left) and the neural network (right). For sake of clarity, only events with a 2 mm step are shown.

size of 3 mm (*I1*) instead of 1 mm (*B1)* for the reference scan slightly improves differential uniformity and distortion on the edges of the FOV. The degradation of the regression when the step scan decreases is probably due to errors in the average light responses. These errors, which could be related to small pedestal or gain fluctuations during measurements, are of the order of magnitude of the expected light variation between two measurement points separated by 1 mm. Therefore, for a set of several consecutive points, the overall trend for the variation in light response is correct but affected by these local fluctuations. The bivariate spline interpolation follows that local fluctuations when using a 1 mm step, whereas the interpolated light response is smoothed with a 3 mm step. The poor estimation of the mean light response at the local scale creates "attractors" during the convergence process, resulting in local heterogeneities.

Although the *J1* configuration uses the parameters that give separately

	Region	SR (mm)	LISD (mm)	DU (%)	IU (%)
Fit method	UFOV	1.44±0.24	0.09±0.12	3.63	5.36
	CFOV	1.30±0.06	0.06±0.03	3.60	4.61
	$\overline{\text{CFOV}}$	1.85±0.77	0.19±0.37	3.72	7.61
Neural network	UFOV	1.31±0.40	0.15±0.11	2.45	4.35
	CFOV	1.15±0.06	0.11±0.05	2.39	3.35
	$\overline{\text{CFOV}}$	1.54±0.55	0.35±0.66	2.47	6.3

Table 3: Summary of the intrinsic spatial performance measured with the best configuration of the fitting method (*J1*) and the neural network (*L2*).

the best spatial performance, some effects are not cumulative. The optimal configuration therefore results from a compromise between the different spatial performances, with a greater weight given to spatial resolution and uniformity. This is why the *J1* configuration is the best in terms of spatial resolution and uniformity, but at the cost of slightly more distortion than the *G1* configuration.

4.2. Neural Network

The number of events in the training dataset is one of the key parameters for good performance of a neural network. For good generalization, the dataset should have as many different examples as possible. Fig. 7 shows that for our regression task, the number of events strongly affects the distortion and thus the spatial uniformity. As for the fitting algorithm, this behavior is asymptotic (*A2* to *C2*) and approximately constant performance is achieved beyond 500 events per position. This asymptotic behavior is due to the fact that above this number, the new added frames are not statis-

tically different from the previous ones, so the learning does not progress significantly. Since we have more events per position available, we chose to use 800 events to push further the homogeneity of the spatial response. The step size of the training dataset, 1 mm or 3 mm (*B2* or *I2*), is also crucial when collecting data. A small step size will allow to give to the network a nearly complete description of the latent space it is trying to map. Thus, the 1 mm step size allows the best performance to be achieved in terms of spatial resolution, distortion and uniformity in the whole FOV (Fig. 7, 8 and 9). All spatial performance improve slightly by adding filters (Fig. 7, 8 and 9). Indeed, the number of filters per layer determines how many features are extracted from the frames (*D2*, *B2* and *E2*). On one hand, if this number is to low, the information contained in the frames is not fully extracted. On the other hand, too many filters per layer adds redundancy and useless complexity, preventing the network to learn quickly and easily. The optimal number of filters per layer is 32, 64, 128 and 256 from the first to the last convolution layer, respectively, which corresponds to the *E2* configuration. The optimal size of the filters to use remains a topic of discussion in artificial neural network research (*F2*, *B2* and *G2*). In our application, 5×5 kernels for all filters (*G2*) works best for low distortion (fig. 8). This can be explained by the local features extracted with 3×3 kernels, such as contours for instance, while the large 5×5 kernels will extract large scale information. The frames we feed the neural network with are sampled continuous charge distributions that contain no local features or very sharp content variations from pixel to pixel. Thus, the 5×5 kernels perform better than the 3×3 kernels for extracting position information [14]. Islam *et al.* [15] also

showed that zero-padding plays a significant role for the good performance of Convolutional Neural Networks extracting absolute position information. The deepness of the neural network, that is the number of *Deep Residual Convolutional Blocks* in our case, will determine how much the extracted information is processed by the network. Like the number of filters per layer, an optimum can be found between a shallow network that does not fully process extracted features, and a too deep network where too much complexity prevents good performance. We see that increasing the number of *Deep Residual Convolutional Blocks* from two to three (*B2* to *H2*) slightly improves all spatial performance, as the extracted features are of higher level (Fig. 7, 8 and 9). We also tested a configuration with only a single residual block, but the spatial performance were poor. Since we kept the number of neurons in the network constant while increasing its deepness, this shows that the complexity of the network (i.e the number of neurons or filters) is not the only parameter to control to reach good performance, but that the deepness of the network is also important to fully process the information contained in the frames. The use of skip connections allows better convergence during the learning phase, and therefore to extract better features from the frames to accurately reconstruct the interaction positions of γ-rays. We see that by removing these shortcuts (*K2*), which leaves us with a conventional Convolutional Neural Networks, performance are not specifically degraded except for the distortion when compared to the same network with residual connections (*B2*) (fig. 8). Indeed, this suppression only results in a global spatial shift by a constant of all the reconstructed positions in the x and y directions. This effect could not be understood. We also show that using the MAE loss

function (*J2*) leads to a better spatial resolution and slightly better overall spatial performance than using MSE loss function (*B2*). We explain this effect by the fact that most of the predictions give errors that are smaller than one from the end of the first epoch. Therefore, as MSE is the squared error and MAE is linear, the loss value will be smaller with MSE and the model will not be penalized as much as with MAE. The drawback of using MAE is that for predictions giving errors larger than one, the model will be less penalized than when using MSE. However, the performance metrics used in our study, that is the spatial resolution FWHM, is not sensitive to outlying values. The numerical issue of the model being less penalized for errors less than one when using MSE could be addressed by multiplying the expected and predicted values by a carefuly chosen constant, so that the error never drops below one. We could therefore limit outlying predictions while getting good predictions when error is below one, as with MAE.

The neural network configuration *L2* gives the best spatial performance of all tested configurations, but as it uses three *Deep Residual Convolutional Blocks* and more filters than all other configurations, its processing speed is slowed down. For instance, its reconstruction rate is approximately divided by 4 compared to the *B2* configuration.

4.3. Comparison of the methods

The fitting algorithm is relatively simple to implement and optimize and gives good overall results with a spatial resolution of 1.30 mm FWHM and a distortion of 0.06 mm in the CFOV, but a poor computational speed of 90 events/s/CPU (table 3). The implementation, tuning and training of basic neural networks is made easy thanks to Keras, but optimization of hyperpa-

rameters was done through empirical method, as the complexity of artificial neural networks only allows to use them as "black boxes". Although systematic optimization of hyperparameters is a heavy process, we obtained a neural network with a spatial resolution of 1.15 mm and a distortion of 0.11 mm FWHM in the CFOV and a suitable reconstruction speed of around 350 events/s/CPU (table 3). Both methods therefore give good spatial performance, but the neural network outperforms the fitting method for spatial rsolution and uniformity, because it is able to learn position information from complex light distributions, whereas for the fitting method, the information is averaged in the mean light response functions of SiPMs. The resulting improvement is even greater near the edges of the detector (table 3), where the light distribution undergoes strong distortions, and should further increase with thicker scintillators. Therefore, thanks to its good performance on the edges of the crystal, the neural network is the algorithm that gives the effective UFOV the closest to the real dimensions of the detection area, which is crucial for small FOV γ-cameras. Those performance are not intrinsic to the reconstruction methods, but embed all the uncertainties due to random physical effects and experimental artefacts. The intrinsic performance of the algorithms are therefore possibly better than the values reported in the result section. The impact of CS+PE events on the reconstruction is the same for the fitting algorithm and the neural network. The slight increase of the interquartile range of prediction error distribution means that those events mainly impact the tails of distributions of reconstructed spots. This could lead to a reduction of image contrast if the contribution of Compton events becomes much larger. Another advantage of the neural network is its recon-

struction speed. Our application requires processing at a minimum rate of around 10 kevents/s. The processing speed of the fitting method is a barrier to its use in a clinical environment, as the image reconstruction will be performed on a clinical computer with limited resources, and the acceleration of this method can only be done through multiprocessing. The neural network, on the other hand, can be speed up using Keras native multithreading implementation, which is a more resource-efficient technique than multiprocessing. To increase further the reconstruction rate of the neural network, it is also possible to slightly reduce its spatial accuracy by limiting the number of layers. In more general conditions, the advent of high performance GPUs are a strong motivation for using neural networks.

5. Conclusion

We developed, optimized and compared two image reconstruction methods for high-resolution γ imaging, both using experimental data for training. The *Deep Residual Convolutional* neural network we have developed fully satisfies the spatial performance targeted for our application in internal radiation therapy, with a mean spatial resolution close to 1 mm and a very low distortion. The processing speed achieved will also allow to use it for real-time applications in a clinical context.

6. Acknowledgments

This work was carried out with the financial support from ITMO Cancer AVIESAN (Alliance Nationale pour les Sicences de la Vie et de la Santé /

National Alliance for Life Sciences & Health) within the framework of the Cancer Plan.

References

[1] C. Trigila et al., "A mobile high-resolution gamma camera for therapeutic dose control during radionuclide therapy", Phys. Med. Biol., Jan. 2022, doi: 10.1088/1361-6560/ac4c31.

[2] R. Pani et al., "Revisited position arithmetics for LaBr3:Ce continuous crystals", Nuclear Physics B - Proceedings Supplements, vol. 197, n 1, p. 383-386, déc. 2009, doi: 10.1016/j.nuclphysbps.2009.10.109.

[3] Z. Li, M. Wedrowski, P. Bruyndonckx, et G. Vandersteen, "Nonlinear least-squares modeling of 3D interaction position in a monolithic scintillator block", Phys. Med. Biol., vol. 55, n 21, p. 6515-6532, nov. 2010, doi: 10.1088/0031-9155/55/21/012.

[4] A. Morozov et al., "Iterative reconstruction of SiPM light response functions in a square-shaped compact gamma camera", Oct. 2016, doi: 10.1088/1361-6560/aa6029.

[5] H. H. Barrett, W. C. J. Hunter, B. W. Miller, S. K. Moore, Y. Chen, et L. R. Furenlid, "Maximum-Likelihood Methods for Processing Signals From Gamma-Ray Detectors", IEEE Transactions on Nuclear Science, vol. 56, n 3, p. 725-735, juin 2009, doi: 10.1109/TNS.2009.2015308.

[6] H. T. van Dam et al., "Improved Nearest Neighbor Methods for Gamma Photon Interaction Position Determination in Monolithic Scintillator

PET Detectors", IEEE Trans. Nucl. Sci., vol. 58, n 5, p. 2139-2147, oct. 2011, doi: 10.1109/TNS.2011.2150762.

[7] "scipy.optimize.least_squares — SciPy v1.9.1 Manual". https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.least_squares.html.

[8] G. Jaliparthi, P. F. Martone, A. V. Stolin, and R. R. Raylman, "Deep residual-convolutional neural networks for event positioning in a monolithic annular PET scanner," Phys. Med. Biol., vol. 66, no. 14, p. 145008, Jul. 2021, doi: 10.1088/1361-6560/ac0d0c.

[9] S. Ioffe et C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift", 2015, doi: 10.48550/ARXIV.1502.03167.

[10] K. He, X. Zhang, S. Ren, et J. Sun, "Deep Residual Learning for Image Recognition", in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, juin 2016, p. 770-778. doi: 10.1109/CVPR.2016.90.

[11] D. P. Kingma et J. Ba, "Adam: A Method for Stochastic Optimization", 2014, doi: 10.48550/ARXIV.1412.6980.

[12] Association, National Electrical Manufacturers 2018 NEMA Standards publication NU 1-2018, Performance measurements of gamma cameras. In: Rosslyn, VA.

[13] B. S. Bhatia, S. L. Bugby, J. E. Lees, et A. C. Perkins, "A scheme for assessing the performance characteristics of small field-of-view gamma

493 cameras", Physica Medica, vol. 31, n 1, p. 98-103, févr. 2015, doi:
494 10.1016/j.ejmp.2014.08.004.

495 [14] M. A. Islam, S. Jia, and N. D. B. Bruce, "How Much Position In-
496 formation Do Convolutional Neural Networks Encode?", 2020, doi:
497 10.48550/ARXIV.2001.08248.

498 [15] M. A. Islam, M. Kowal, S. Jia, K. G. Derpanis, et N. D. B. Bruce, "Posi-
499 tion, Padding and Predictions: A Deeper Look at Position Information
500 in CNNs", 2021, doi: 10.48550/ARXIV.2101.12322.