



HAL
open science

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-04446415>

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

1 Optimized Reconstruction of the Position of Interaction 2 in High-Performances γ -Cameras

3 T. Bossis^{a,b}, M-A. Verdier^{a,b}, L. Pinot^{a,b}, F. Bouvet^{a,b}, T. Beaumont^c,
4 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,*

5 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 cm² field of view dedicated to thyroid imaging during diseases treatments with ¹³¹I. 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 submillimeter distortion in

the center of the field of view.

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

8 **1. Introduction**

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

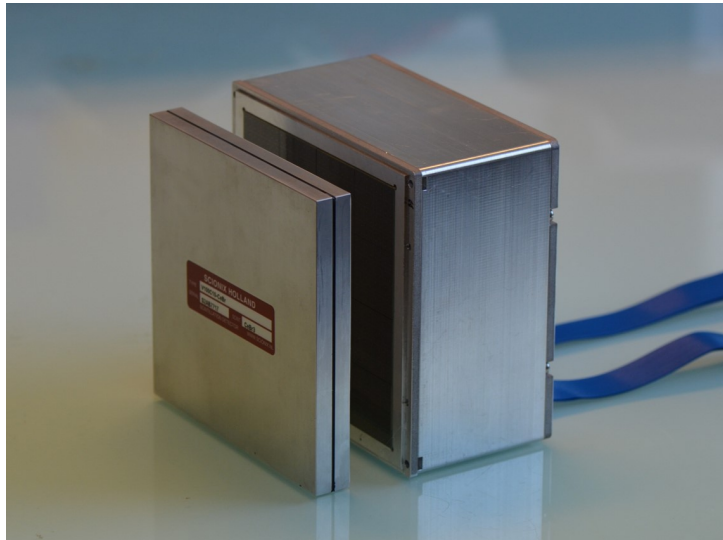


Figure 1: Photodetection module of the mobile camera composed of a 256 SiPMs array coupled to an encapsulated $10 \times 10 \times 1 \text{ cm}^3$ CeBr_3 monolithic scintillator.

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

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

56 **2. Material and methods**

57 *2.1. Global design of the γ -detection system*

58 The γ -detection system of the mobile camera is composed of a $10 \times 10 \text{ cm}^2$
59 and 1 cm thick CeBr_3 monolithic scintillator (Scionix) encapsulated in a me-
60 chanical frame with a 1.5 mm thick glass output window and a reflective op-
61 tical coating. The scintillator is coupled with optical grease (BC-630, Saint
62 Gobain) to a 4×4 Hamamatsu S13361-6050 arrays of 4×4 silicon photomulti-
63 pliers (SiPMs). The pixels have a sensitive area of $6 \times 6 \text{ mm}^2$ and a microcell
64 pitch of $50 \mu\text{m}$. The SiPMs signals are readout by commercial acquisition
65 electronics manufactured by PETSys Electronics. This electronic readout

66 consists of four TOFPET 2B ASICs with 64 analog readout channels each,
67 used in a charge-integration mode and connected to a DAQ computer via a
68 Gigabit Ethernet link. The SiPMs bias and electronic triggering threshold
69 were adjusted to optimize the energy resolution of the camera. The over-
70 voltage was set to 7.5 V, which corresponds to a gain of $4.3 \cdot 10^6$ and a PDE
71 of 55%. Thermal noise events (11 kcps for the optimal operative bias and
72 individual trigger threshold used) are completely suppressed by an internal
73 hardware trigger (10 ns coincidence window), which operates between two
74 regions of the detector defined as a checkerboard, where each square cor-
75 responds to a 4×4 SiPM array. The non-uniformity of the photodetector
76 light response was evaluated and corrected by irradiating the SiPMs with a
77 pulsed LED source. The relative standard deviations of the light response
78 over the 256 SiPMs before and after uniformity correction are 12.8% and
79 0.96%, respectively.

80 *2.2. Experimental Data*

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

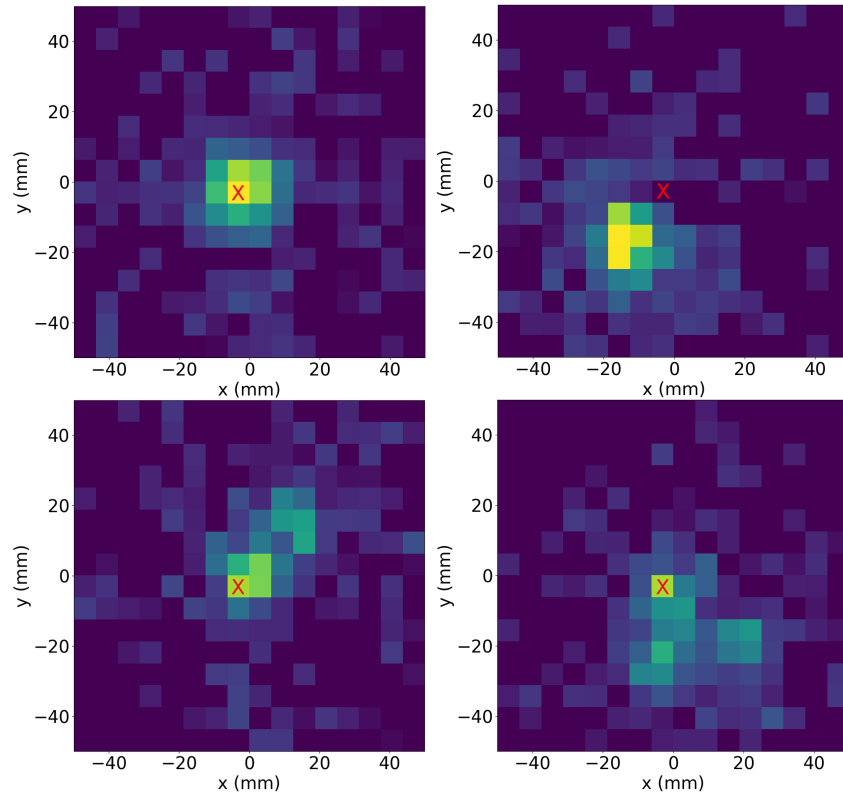


Figure 2: Typical charge distributions produced after the total energy absorption of a 364 keV γ -ray in the monolithic scintillator : (Top-left) photoelectric effect after passing through the hole of the collimator, (Top-right) photoelectric effect after penetrating through the collimator material and (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.

91 reconstruction algorithms if good spatial performance is to be achieved. A
92 first energy selection of twice the energy resolution was applied around the
93 356 keV emission of ^{133}Ba (16% energy window) to only keep events corre-
94 sponding to a full deposition of the incident energy photon. Fig. 2 shows
95 typical patterns of the frames produced by the full energy absorption of a
96 356 keV γ -ray in a 1 cm thick CeBr_3 scintillator. The scan dataset was then
97 filtered by a second spatial selection to remove full-energy events that have
98 penetrated through the collimator material (fig. 2 top-right) or that have
99 scattered once or several times by Compton interaction inside the scintilla-
100 tor, before being absorbed by photoelectric interaction, which will be called
101 CS+PE events in the rest of the manuscript for concision (fig. 2 bottom left
102 and right). The penetration events represent 77% of all events detected in
103 the energy window, while CS+PE events remain negligible with less than
104 1% of all detected events. The 22% remaining events correspond to 356 keV
105 gamma rays that propagated through the collimator hole and were fully ab-
106 sorbed inside the scintillator after a single photoelectric interaction (fig. 2,
107 top-left). Both types of spurious events do not produce a well-localized en-
108 ergy deposition and therefore result in poor localization. In order to clean
109 data from parasite events, the position of all events was reconstructed with
110 a non-linear weighting method [2] and only those located within a radius of
111 3 mm around the mean reconstructed position were kept (the value of the
112 radius was empirically optimized). Two thousand full-energy interactions
113 were then obtained for each scan position after spatial filtering of spurious
114 events. Half of these events were used as a reference dataset for training of
115 the reconstruction methods, while the other half were used as a test dataset

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

123 *2.3. Reconstruction algorithms*

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

127 *2.3.1. Interpolated model fitting*

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

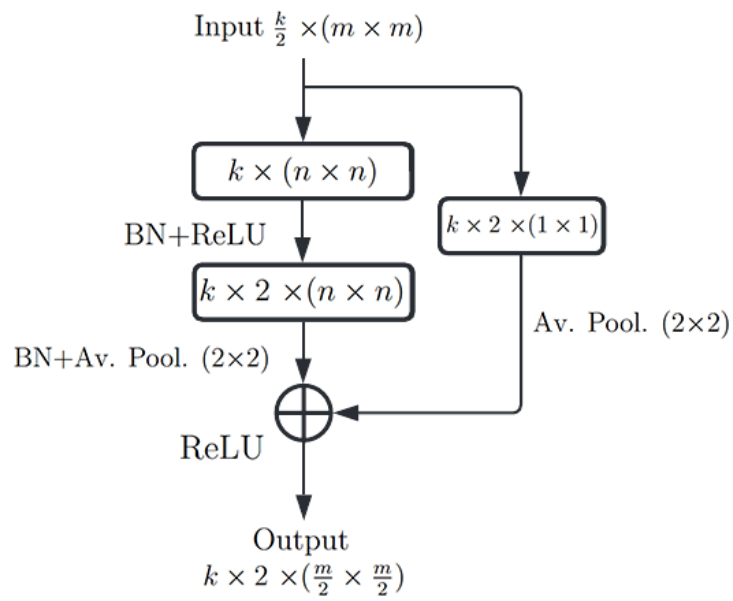


Figure 3: Structure of a *Deep Residual Convolutional Block*. m is the dimension of the input image, n is the dimension of the filters and k is the number of filters. *BN* stands for Batch Normalization, *Av.Pool.* for Average Pooling and *ReLU* for Rectified Linear Unit.

140 (3×3, 5×5, ...). The trade-off between precision and convergence speed of
141 the fitter can also be optimized by changing its tolerance for termination
142 parameter [7].

143 2.3.2. Neural network

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

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

176 2.4. Performance metrics

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

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

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

210 *2.5. Optimization of the algorithms*

211 The different tested configurations for the fitting method and the neural
212 network are summarized respectively in table 1 and 2. Each time a parameter
213 is changing, all other parameters remain constant and identical to those of
214 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.

215 network. The optimized parameters for the fitting algorithm are the number
216 of events/position of the reference scan, the step of the reference scan, the
217 size of the gaussian smoothing kernel and the convergence tolerance of the
218 fitter [7]. The optimized parameters for the neural network are the number of
219 events/position and the step of the scan for the training dataset, the number
220 of filters per layer, the number of residual blocks, the loss function and the
221 effect of the residual connection. The model has been trained for 30 epochs
222 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.

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

229 performances are the mean of the three trained models.

230 **3. Results**

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

243 *3.1. Fitting algorithm*

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

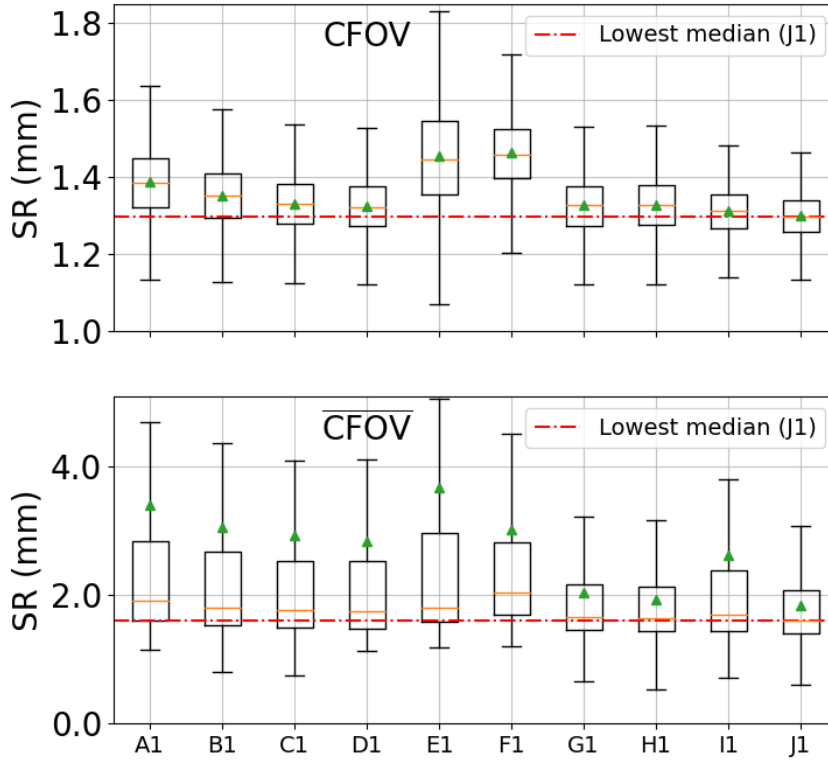


Figure 4: Box plots of the spatial resolution distributions (in mm FWHM) for the different configurations of the fit method. Top: Spatial resolution in CFOV. Bottom: Spatial resolution in $\overline{\text{CFOV}}$.

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

257 3.2. Neural Network

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

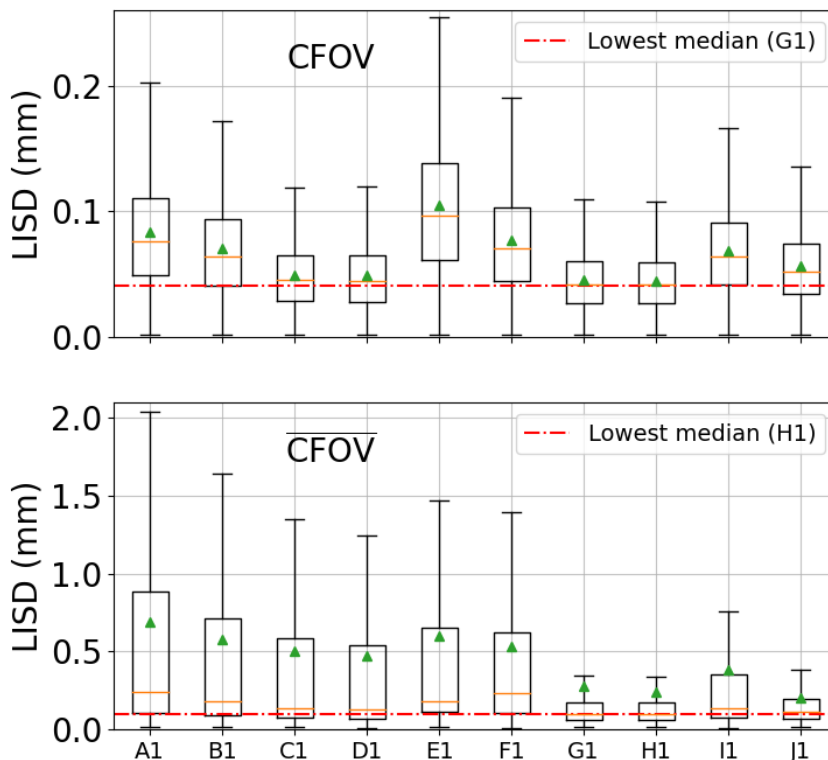


Figure 5: Box plots of the LISD value distributions (in mm) for the different configurations of the fit method. Top: Distortion in CFOV. Bottom: Distortion in $\overline{\text{CFOV}}$.

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

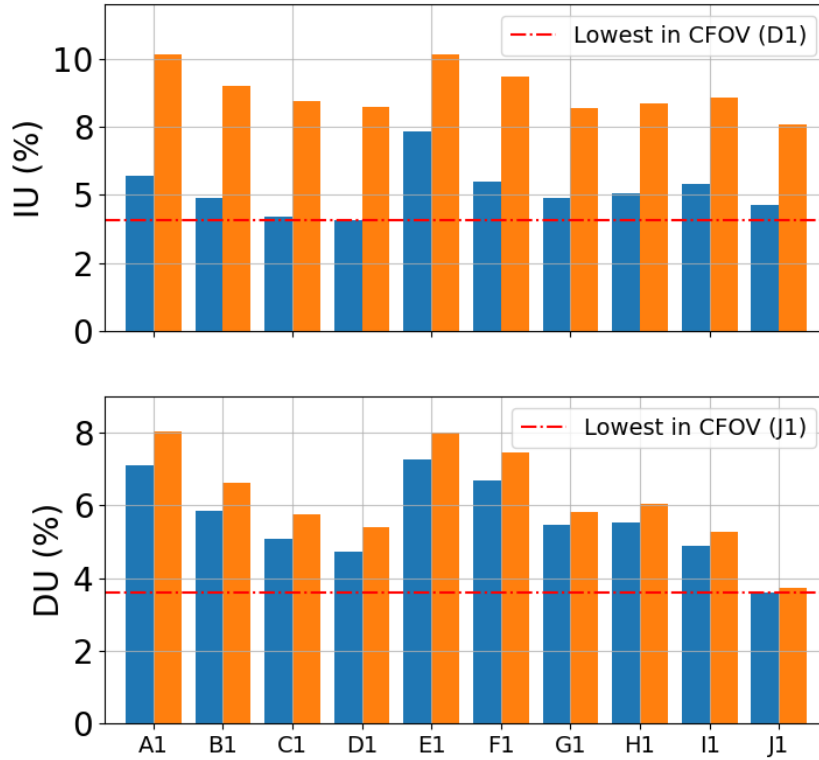


Figure 6: Bar graphs of differential and integral uniformity (in %) for the different parameters configurations of the fitting algorithm. Blue: Uniformity in CFOV. Orange: Uniformity in the $\overline{\text{CFOV}}$.

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

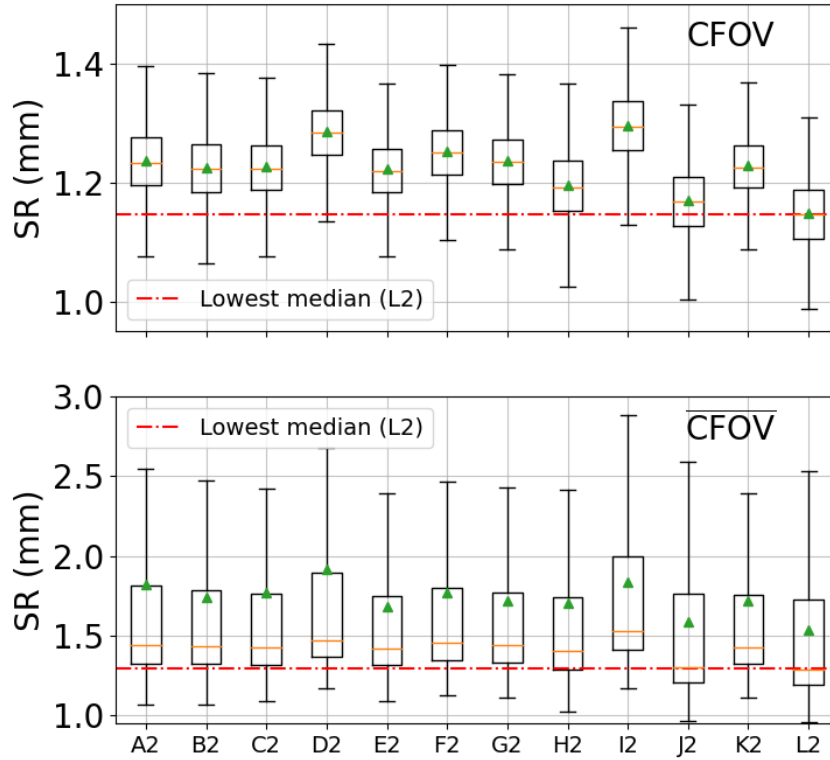


Figure 7: Box plots of the spatial resolution distributions (in mm FWHM) for the different hyperparameters configurations of the neural network. Top: Spatial resolution in CFOV. Bottom: Spatial resolution in $\overline{\text{CFOV}}$.

273 4. Discussion

274 4.1. Fitting algorithm

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

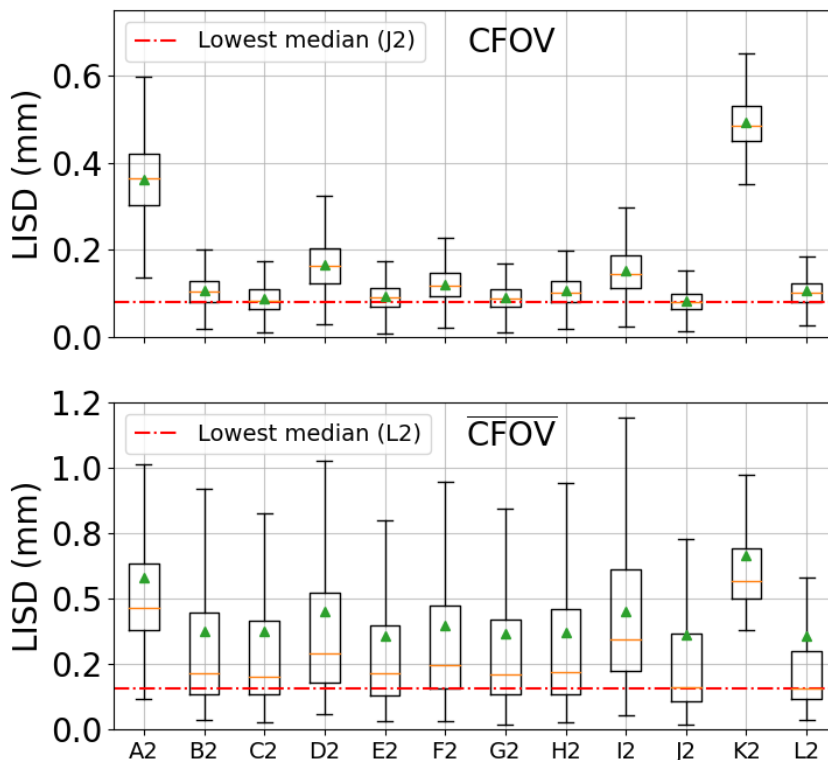


Figure 8: Box plots of the LISD value distributions (in mm) for the different hyperparameters configurations of the neural network. Top: Distortion in CFOV. Bottom: Distortion in $\overline{\text{CFOV}}$.

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

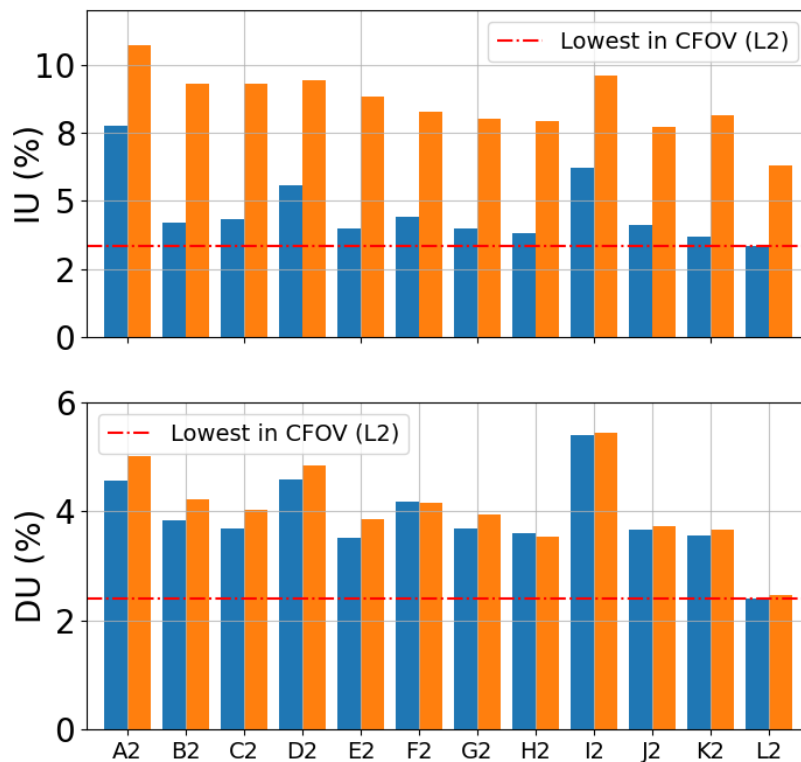


Figure 9: Bar graphs of differential and integral uniformity (in %) for the different hyperparameters configurations of the neural network. Blue: Uniformity in CFOV. Orange: Uniformity in the $\overline{\text{CFOV}}$.

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

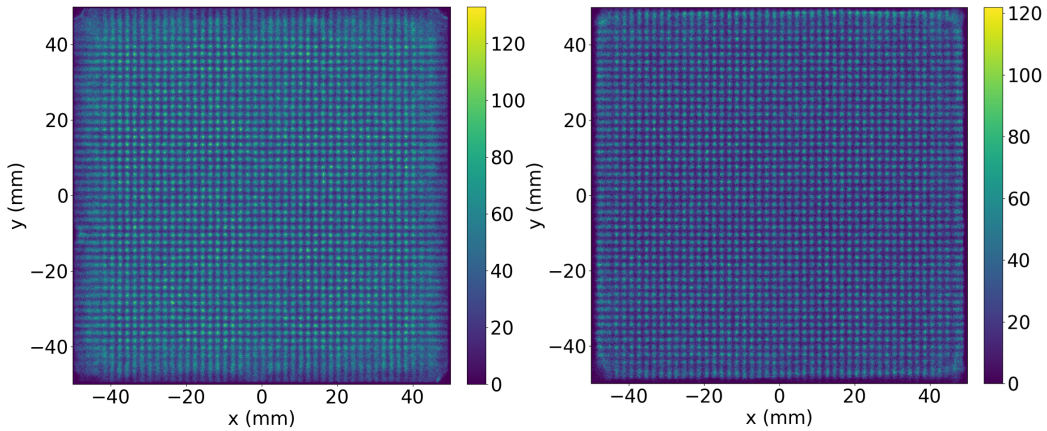


Figure 10: Images of a reconstructed scan obtained with the optimized configurations of the fitting algorithm (left) and the neural network (right). For sake of clarity, only events with a 2 mm step are shown.

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

310 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 ($L2$).

311 the best spatial performance, some effects are not cumulative. The opti-
312 mal configuration therefore results from a compromise between the different
313 spatial performances, with a greater weight given to spatial resolution and
314 uniformity. This is why the $J1$ configuration is the best in terms of spatial
315 resolution and uniformity, but at the cost of slightly more distortion than
316 the $G1$ configuration.

317 4.2. Neural Network

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

326 tically different from the previous ones, so the learning does not progress
327 significantly. Since we have more events per position available, we chose to
328 use 800 events to push further the homogeneity of the spatial response. The
329 step size of the training dataset, 1 mm or 3 mm (*B2* or *I2*), is also crucial
330 when collecting data. A small step size will allow to give to the network a
331 nearly complete description of the latent space it is trying to map. Thus,
332 the 1 mm step size allows the best performance to be achieved in terms of
333 spatial resolution, distortion and uniformity in the whole FOV (Fig. 7, 8
334 and 9). All spatial performance improve slightly by adding filters (Fig. 7, 8
335 and 9). Indeed, the number of filters per layer determines how many features
336 are extracted from the frames (*D2*, *B2* and *E2*). On one hand, if this num-
337 ber is too low, the information contained in the frames is not fully extracted.
338 On the other hand, too many filters per layer adds redundancy and useless
339 complexity, preventing the network to learn quickly and easily. The optimal
340 number of filters per layer is 32, 64, 128 and 256 from the first to the last
341 convolution layer, respectively, which corresponds to the *E2* configuration.
342 The optimal size of the filters to use remains a topic of discussion in arti-
343 ficial neural network research (*F2*, *B2* and *G2*). In our application, 5×5
344 kernels for all filters (*G2*) works best for low distortion (fig. 8). This can be
345 explained by the local features extracted with 3×3 kernels, such as contours
346 for instance, while the large 5×5 kernels will extract large scale informa-
347 tion. The frames we feed the neural network with are sampled continuous
348 charge distributions that contain no local features or very sharp content vari-
349 ations from pixel to pixel. Thus, the 5×5 kernels perform better than the
350 3×3 kernels for extracting position information [14]. Islam *et al.* [15] also

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

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

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

395 4.3. Comparison of the methods

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

401 rameters was done through empirical method, as the complexity of artificial
402 neural networks only allows to use them as "black boxes". Although sys-
403 tematic optimization of hyperparameters is a heavy process, we obtained
404 a neural network with a spatial resolution of 1.15 mm and a distortion of
405 0.11 mm FWHM in the CFOV and a suitable reconstruction speed of around
406 350 events/s/CPU (table 3). Both methods therefore give good spatial per-
407 formance, but the neural network outperforms the fitting method for spatial
408 resolution and uniformity, because it is able to learn position information from
409 complex light distributions, whereas for the fitting method, the information
410 is averaged in the mean light response functions of SiPMs. The resulting im-
411 provement is even greater near the edges of the detector (table 3), where the
412 light distribution undergoes strong distortions, and should further increase
413 with thicker scintillators. Therefore, thanks to its good performance on the
414 edges of the crystal, the neural network is the algorithm that gives the ef-
415 fective UFOV the closest to the real dimensions of the detection area, which
416 is crucial for small FOV γ -cameras. Those performance are not intrinsic to
417 the reconstruction methods, but embed all the uncertainties due to random
418 physical effects and experimental artefacts. The intrinsic performance of the
419 algorithms are therefore possibly better than the values reported in the re-
420 sult section. The impact of CS+PE events on the reconstruction is the same
421 for the fitting algorithm and the neural network. The slight increase of the
422 interquartile range of prediction error distribution means that those events
423 mainly impact the tails of distributions of reconstructed spots. This could
424 lead to a reduction of image contrast if the contribution of Compton events
425 becomes much larger. Another advantage of the neural network is its recon-

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

437 **5. Conclusion**

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

445 **6. Acknowledgments**

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

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

450 **References**

451 [1] C. Trigila et al., “A mobile high-resolution gamma camera for therapeutic
452 dose control during radionuclide therapy”, *Phys. Med. Biol.*, Jan.
453 2022, doi: 10.1088/1361-6560/ac4c31.

454 [2] R. Pani et al., “Revisited position arithmetics for LaBr₃:Ce continuous
455 crystals”, *Nuclear Physics B - Proceedings Supplements*, vol. 197, n 1,
456 p. 383-386, déc. 2009, doi: 10.1016/j.nuclphysbps.2009.10.109.

457 [3] Z. Li, M. Wedrowski, P. Bruyndonckx, et G. Vandersteen, “Nonlinear
458 least-squares modeling of 3D interaction position in a monolithic scin-
459 tillator block”, *Phys. Med. Biol.*, vol. 55, n 21, p. 6515-6532, nov. 2010,
460 doi: 10.1088/0031-9155/55/21/012.

461 [4] A. Morozov et al., “Iterative reconstruction of SiPM light response
462 functions in a square-shaped compact gamma camera”, Oct. 2016, doi:
463 10.1088/1361-6560/aa6029.

464 [5] H. H. Barrett, W. C. J. Hunter, B. W. Miller, S. K. Moore, Y. Chen, et
465 L. R. Furenlid, “Maximum-Likelihood Methods for Processing Signals
466 From Gamma-Ray Detectors”, *IEEE Transactions on Nuclear Science*,
467 vol. 56, n 3, p. 725-735, juin 2009, doi: 10.1109/TNS.2009.2015308.

468 [6] H. T. van Dam et al., “Improved Nearest Neighbor Methods for Gamma
469 Photon Interaction Position Determination in Monolithic Scintillator

- 470 PET Detectors”, IEEE Trans. Nucl. Sci., vol. 58, n 5, p. 2139-2147, oct.
471 2011, doi: 10.1109/TNS.2011.2150762.
- 472 [7] “scipy.optimize.least_squares — SciPy v1.9.1 Manual”. https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.least_squares.html.
473
474
- 475 [8] G. Jaliparthi, P. F. Martone, A. V. Stolin, and R. R. Raylman, “Deep
476 residual-convolutional neural networks for event positioning in a mono-
477 lithic annular PET scanner,” Phys. Med. Biol., vol. 66, no. 14, p. 145008,
478 Jul. 2021, doi: 10.1088/1361-6560/ac0d0c.
- 479 [9] S. Ioffe et C. Szegedy, “Batch Normalization: Accelerating Deep
480 Network Training by Reducing Internal Covariate Shift”, 2015, doi:
481 10.48550/ARXIV.1502.03167.
- 482 [10] K. He, X. Zhang, S. Ren, et J. Sun, “Deep Residual Learning for Image
483 Recognition”, in 2016 IEEE Conference on Computer Vision and Pat-
484 tern Recognition (CVPR), Las Vegas, NV, USA, juin 2016, p. 770-778.
485 doi: 10.1109/CVPR.2016.90.
- 486 [11] D. P. Kingma et J. Ba, “Adam: A Method for Stochastic Optimization”,
487 2014, doi: 10.48550/ARXIV.1412.6980.
- 488 [12] Association, National Electrical Manufacturers 2018 NEMA Standards
489 publication NU 1-2018, Performance measurements of gamma cameras.
490 In: Rosslyn, VA.
- 491 [13] B. S. Bhatia, S. L. Bugby, J. E. Lees, et A. C. Perkins, “A scheme for
492 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.