Convolutional networks lead to large improvements in time estimation for monolithic PET detectors

1st Jens Maebe
Department of Electronics and Information Systems
Ghent University
Ghent, Belgium
jens.maebe@ugent.be

2nd Stefaan Vandenberghe
Department of Electronics and Information Systems
Ghent University
Ghent, Belgium
stefaan.vandenberghe@ugent.be

***Abstract*—We investigate the use of convolutional neural networks for time estimation in monolithic scintillation crystals. The required data is obtained by Monte Carlo simulation in GATE v8.2, using a monolithic LYSO crystal coupled to a readout array of silicon photomultipliers (SiPMs). The electronic signals are then simulated as a sum of bi-exponentional functions centered around the scintillation photon detection times, including noise generation from various sources. Two network architectures are explored, a 2D convolutional neural network (CNN) with as input the SiPM timestamps obtained by leading edge discrimination, and a 3D CNN with as input the detector waveforms themselves. The 2D network showed a single detector time resolution of 106 ps full width at half maximum (FWHM), improving performance by 16% compared to the 126 ps FWHM obtained from a simple averaging of the first three recorded timestamps. The 3D network offered a time resolution of 88 ps FWHM, a 30% improvement. These advances can lead to higher signal-to-noise ratios in time-of-flight positron emission tomography, ultimately resulting in better diagnostic capabilities.**

***Index Terms*—PET, time-of-flight, scintillation detectors, deep learning**

I. INTRODUCTION

Monolithic scintillation crystals have been proposed for gamma detection in positron emission tomography (PET) due to their cost-effectiveness combined with high spatial resolution and depth-of-interaction decoding capabilities. Time estimation however remains difficult as the scintillation light is spread out over multiple silicon photomultiplier (SiPM) pixels, negatively impacting the signal-to-noise ratio in time-of-flight PET. Traditionally, a timestamp is generated for each SiPM by e.g. leading edge discrimination on the electronic signal, after which the gamma impinging time is estimated by an averaging of the first few recorded timestamps. Alternatively, maximum likelihood interaction time estimation (MLITE) [1] offers improved time resolution but requires training data with known gamma interaction positions for building lookup-tables. Inspired by the success of deep learning approaches for time-of-flight estimation between pixelated detectors [2], we perform a simulation study to investigate convolutional neural networks (CNNs) for time estimation in monolithic scintillation detectors. Either the SiPM timestamps obtained by leading edge discrimination or the original waveforms themselves are used as input.

II. MATERIALS AND METHODS

A. Data Generation

GATE v8.2 is used for Monte Carlo simulation of 511 keV gamma photons impinging on a monolithic 50x50x16 mm^3 LYSO crystal. An 8x8 readout array of 6.07x6.07 mm^2 SiPMs based on the sensL J-Series detects the generated scintillation light. The detection times t_p of scintillation photons p are then used to generate the SiPM response s as a sum of bi-exponential functions with rise-time τ_r = 114 ps and fall-time τ_f = 50 ns:

$$s(t) = \sum_p \left(\exp\left[\frac{t_p - t}{\tau_f}\right] - \exp\left[\frac{t_p - t}{\tau_r}\right] \right) H(t - t_p) \quad (1)$$

where H denotes the Heaviside function. The simulations include a limited photon detection efficiency of 50%, a dark count rate of 150 kHz/mm^2, a Gaussian time jitter between photon arrival at the readout surface and generation of the output pulse with a standard deviation of 50 ps, optical crosstalk with a probability of 25% and white Gaussian electronic noise with standard deviation equal to 10% of a single photoelectron pulse. A detailed description of the signal generation process can be found in [3].

Only taking into account events that deposit their full energy in the crystal, a dataset of 10 million gamma photons uniformly and perpendicularly incident on the front surface of the crystal is obtained. It is further subdivided into a training, validation and test set of 6, 2 and 2 million events respectively. Making use of the 8-fold symmetry in the crystal, only 12.5% of each dataset needs to be simulated.

B. Network Training

Two network architectures are investigated. The first is a 2D CNN (kernel size 3x3) using the 8x8 array of SiPM timestamps as input, obtained by leading edge discrimination at the 1 photoelectron level after baseline correction. The timestamps are offset so that the first always occurs at $t = 0$, and passed through a tanh function in order to maintain high granularity for early timestamps while not discarding later ones. The second network is a 3D CNN (kernel size 3x3x5) using an 8x8x30 array as input, corresponding to a 3 ns window (100 ps binning) of the rising edge of the

This study was funded by Research Foundation Flanders (FWO).

SiPM signals. The window starts 500 ps prior to the onset of the sum of all 64 SiPM signals. Both networks consist of 3 convolutional layer blocks followed by 2 fully connected layer blocks, each block except the last containing a batch normalization layer and ReLU activation function. A single output neuron predicts the gamma impinging time, either with respect to the first timestamp, or with respect to the start of the 3 ns window. Prior to training, the true gamma impinging times of the training and validation set are convolved with a Gaussian to simulate the use of a reference detector with a time resolution of 100 ps full width at half maximum (FWHM). Training is performed with the Adam optimization algorithm, using a learning rate of 0.001 and mini-batch size of 256. The mean squared error between reference and predicted gamma impinging time is used as the loss function, stopping training when the validation loss does not improve for 10 epochs.

III. RESULTS AND DISCUSSION

We investigate the error distribution between the true and predicted gamma impinging times of the test set and compare with a simple averaging of the first three timestamps and the results predicted by MLITE. Since the error distribution is not well described by a Gaussian, we report the standard deviation σ in addition to the FWHM and full width at tenth maximum (FWTM) obtained from a kernel density estimation fit.

An overview of the results is given in table I. Both the 2D and 3D CNN outperform the other methods, improving the FWHM by 16% and 30% respectively when compared to the averaging method. We notice that the time resolution improves slightly when limiting ourselves to events impinging on the detector center. This can also be seen in Fig. 1, showing σ as a function of impinging position along the lateral direction. The oscillations observed in the averaging method can be attributed to a decreased time resolution when gamma interactions occur on top of an SiPM pixel, limiting most of the useful information to a single timestamp. This effect is removed for the neural network methods, although the degradation at the edges remains. Similarly, Fig. 2 shows σ as a function of the depth of the first gamma interaction. Interactions close to the SiPM surface again lead to a degradation in time resolution for the averaging method as the scintillation light is less spread out, resulting in fewer useful timestamps. This effect is less pronounced in the other approaches.

In addition to the improved performance, the 2D CNN offers two more advantages compared to MLITE. Whereas MLITE requires known gamma interaction positions for building the lookup-tables, no such information is required for the neural network approach, simplifying the process of obtaining experimental training data. Inference is also significantly faster, possibly allowing for "online" time estimation using onboard electronics. The 3D CNN offers further improvements to the time resolution by having access to the original signals, rather than replacing this complex but useful information by a single estimator (i.e. leading edge discrimination) prior to training.

TABLE I
GAMMA IMPINGING TIME PREDICTION ERRORS.

Method	Whole Detector			Detector Center[a]		
	σ	FWHM	FWTM	σ	FWHM	FWTM
average	64	126	250	61	122	238
MLITE	56	119	224	56	116	221
2D CNN	51	106	204	48	103	195
3D CNN	41	88	168	38	83	159

Results in picoseconds

[a] Central 20x20 mm^2 of the detector

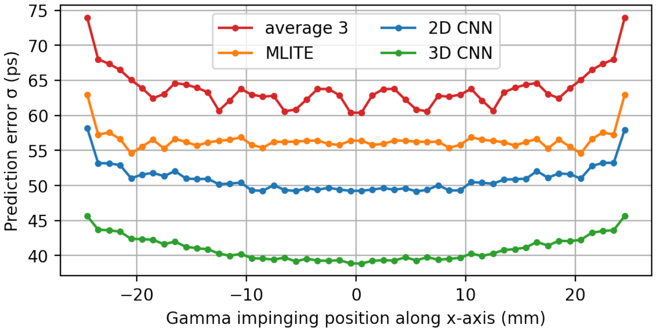

Fig. 1. Gamma impinging time prediction error σ along the lateral direction.

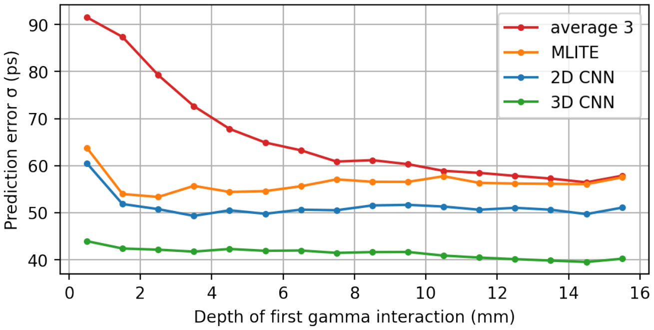

Fig. 2. Gamma impinging time prediction error σ in function of first gamma interaction depth. The SiPM surface is located at 16 mm.

IV. CONCLUSION

In this work, we used simulation data to test a 2D and 3D CNN for time estimation in monolithic PET detectors. The neural network approaches outperformed both a simple averaging of the first three timestamps as well as the maximum likelihood based method MLITE, showing the promising capabilities of convolutional networks for time-of-flight PET.

REFERENCES

[1] H. T. van Dam, G. Borghi, S. Seifert, and D. R. Schaart, "Sub-200 ps CRT in monolithic scintillator PET detectors using digital SiPM arrays and maximum likelihood interaction time estimation," Phys. Med. Biol., vol. 58, no. 10, pp. 3243–3257, Apr. 2013.

[2] E. Berg and S. R. Cherry, "Using convolutional neural networks to estimate time-of-flight from PET detector waveforms," Phys. Med. Biol., vol. 63, no. 2, p. 02LT01, Jan. 2018.

[3] F. Acerbi and S. Gundacker, "Understanding and simulating SiPMs," Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, vol. 926, pp. 16–35, May 2019.