2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Deep learning based timing calibration for PET

Huai Chen and Huafeng Liu*

Abstract— **Neural network has been found an increasingly wide utilization in all fields. Owing to the fact that the traditional optimized algorithm, Iterative Shrinkage-Thresholding Algorithm (ISTA) or Alternating Direction Method of Multipliers (ADMM), could be presented by a form of network, and it could overcome some shortcomings of traditional algorithms, which inspired us to introduce the structured deep network into PET timing calibration. In this paper, by reformulating an ADMM algorithm to a deep network, we introduce a ADMM-Net framework for calibration, which combines the advantage of compatibility of consistency condition method. To verify the performance, several experiments of Monte Carlo simulation in GATE are performed.**

I. INTRODUCTION

Positron Emission Tomography (PET) owns the ability to pair two 511 keV photons in a narrow time window, which indicates the importance of accurate timing information. The width of the window is determined by the longest useful line-of-response (LOR) and the speed of light plus the system's coincidence timing resolution. The timing calibration methods could be categorized as direct and indirect methods. The typical approach of direct calibration is to use a time reference probe. [1], [2]. While the indirect methods for calibration are diversity, some methods are based on iterative optimization in block level [3], or in crystal level [4]. Partial calibration is worked by measuring a known activity distribution such like a point source [1], a rotating line source[5], or a shell [6]. Werner, et al. presented that the scan data of patients or phantoms could be used for calibration [7]. Defrise proposed a new data driven approach in 2D time-of-flight (TOF) PET for calibration, which was derived from the consistency equation [8]. On the basis of that, Li extended this method for 3D TOF PET [9]. Some methods make use of the framework of linear formulation to calculate the offset, the linear system matrix can be established by numbers of LORs. The methods calculated offset by solving linear least squares [10], [11], or added the regularization by penalized least squares estimation [12], [13]. Alternatively, the linear system matrix built by numbers of coincidence events, Reynolds formulated the maximum-likelihood problem amounts to solve the timing calibration problem [14]. Following Reynolds' framework, Freese proposed that the LSQR [15] or LSMR [16] algorithm to accelerate calculation because of the sparsity nature of system matrix, and they applied l_1 minimization instead of least squares estimation to improve the robustness to outliers [17]. For other purpose, there is some method that makes use of regularization to calculate constant offset [18].

Recently, deep neural network has been an effective tool. The traditional algorithms are presented by the form of network [19], [20], and this framework could overcome some shortcomings of traditional methods in calibration, such as long acquisition time, long computation time and other problems.

In this paper, we introduce the ADMM-Net by reformulating an Alternating Direction Method of Multipliers (ADMM) algorithm into network frame for calibration. For improving compatibility, the label of network is using the result of the data driven method derived from the consistency condition. And the deep network framework combines the advantages of ADMM algorithm and the compatibility of data consistency. To assess the performance of proposed method, several experiments are performed by Monte Carlo simulations in GATE.

II. METHOD

A. Linear formulation

Considering that there is a list m of coincidence events data stream, we frame the delay problem as a least squares problem, also referred to as l_2-norm minimization. The problem then could be stated as:

$$\text{minimize } \|Ax-b\|_2^2 \tag{1}$$

where A is system matrix, x is offset vector, and b is measured time difference vector. System matrix A is a $m \times n$ matrix, m is the number of coincidence events, n is the number of detector units. Each row of A has a 1 and -1, and other place is 0, the position of 1 and -1 presented the index of crystal unit, and -1 indicates that it is subtracted from the time of index of 1 crystal unit. Then the corresponding to the row of b is the time difference corresponding to the row of A. A simple example modeling matrix A for a system with four crystal elements and three coincidence events is presented in Eqn.(2).

$$\|Ax-b\|_2^2 = \left\| \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} - \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \right\|_2^2 \tag{2}$$

As mentioned above, A is a sparse matrix, the calibration process could be accelerated by LSQR algorithm. Yet, compared with the l_2-norm, the l_1-norm could obtain a more accurate estimation. Several techniques could potentially improve the stability in presence of noise, such as replace the least squares estimation by l_1 minimization [17], or add l_1 constraint to equation [12]. Both of these problem could be

Huai Chen and Huafeng Liu are with State Key Laboratory of Modern Optical Instrumentation, Zhejiang University, Hangzhou 310027
*Corresponding author (Email:liuhf@zju.edu.cn)

solved by the ADMM algorithm [21], [22]. In this paper, we select adding l_1 regularized loss minimization in to Eqn.(2), and the loss function multiply the 1/2. Then the problem is redefined as an optimization problem as follows:

$$\text{minimize} \quad (1/2)\|Ax-b\|_2^2+\lambda\|x\|_1 \tag{3}$$

this form is l_1 regularized linear regression, also called the lasso [23]. In ADMM form, the lasso problem can be written as

$$\begin{aligned} &\text{minimize } f(x)+g(z) \\ &\text{subject to } x-z=0 \end{aligned} \tag{4}$$

where $f(x)=(1/2)\|Ax-b\|_2^2$ and $g(z)=\lambda\|z\|_1$, and according to the [21], the problem can be written:

$$\begin{aligned} x^{k+1} &:= \left(A^TA+\rho I\right)^{-1}\left(A^Tb+\rho\left(z^k-u^k\right)\right) \\ z^{k+1} &:= S_{\lambda/\rho}\left(x^{k+1}+u^k\right) \\ u^{k+1} &:= u^k+x^{k+1}-z^{k+1} \end{aligned} \tag{5}$$

where the I is the identity matrix, ρ is the augmented Lagrangian parameter, u^k is defined as the running sum of residuals. And soft thresholding operator S is defined as

$$S_\kappa(a)=\begin{cases} a-\kappa & a>\kappa \\ 0 & |a|\le\kappa \\ a+\kappa & a<-\kappa \end{cases} \tag{6}$$

B. Consistency condition

We follow the Defrise' the notation and the discretization process of their work [8], then calculate the offset, since this method could be applied to data measured with an arbitrary tracer distribution, which improves the compatibility in different situation.

Assuming that the circular ring scanner of radius R with $2N$ equally spaced detectors, and the measured data as m_{j,k,i_t} with TOF sampling Δ_t:

$$\begin{aligned} \alpha_j &= j\pi/N \quad j=0,\dots,2N-1 \\ \beta_{j,k} &= \alpha_j+\pi+k\pi/N \quad -k_{\max}\leqslant k\leqslant k_{\max} \\ t &= i_t\Delta_t \quad i_t=-i_m,\dots,i_m \end{aligned} \tag{7}$$

where α, β is the angular coordinates of the two detectors in coincidence, j is the index number of detector, $\beta_{j,k}$ is the fan of detectors in coincidence with α_j, $k=-N+(j_2-j_1)\,\%(2N)$ is the fan index of (j_1,j_2).

The two parameters, data moments $M_{0,j,k}=\sum_{i_t}m_{j,k,i_t}$ and $M_{1,j,k}=\Delta_t\sum_{i_t}i_tm_{j,k,i_t}$, are related to $2N\times 2N$ maxtrix U and vector Y:

$$U_{j_1,j_2}=T_{j_1}\delta_{j_1,j_2}-\Delta_{j_2}M_{0,j_1,k} \tag{8}$$

$$Y_{j_1}=\sum_{k=-k_{\max}}^{k_{\max}}\Delta_{j_2}\left(M_{1,j_1,k}+R\left|\sin\left(\frac{\pi(N+k)}{2N}\right)\right|M_{0,j_1,k}\right) \tag{9}$$

where δ is the Kronecker delta, the diagonal term is $T_{j_1}=\sum_k\Delta_{j_2}M_{0,j_1,k}$, and the Δ_j is weight of detector j at α_j.

To avoid indeterminacy, they added a diagonal matrix $\mu\lambda_{\max}I$ to U in their work, with I is the identity matrix, μ is the trade off parameter and λ_{max} the largest eigenvalue of U. Then we solve the linear equation $(U+\mu\lambda_{\max}I)\cdot x=Y$ for x by gaussian elimination. The calculated x is regarded as the label for network.

C. ADMM-Net

By taking full advantage of the merits of ADMM based methods, the basic idea of ADMM-Net is to map the previous ADMM update steps to a deep network architecture. In Yan' work, they extent and generalize iteration in Eqn.(5), they generalize four types of operations to have learnable parameter as network layers. These operations are generalized as reconstruction layer, convolution layer, non-linear transform layer, and multiplier update layer. In this paper, the convolution layer is not necessary, so we eliminate the convolution layer, remaining the rest three layer in our work.

Reconstruction layer ($\mathbf{X^{(n)}}$): This layer is following the reconstruction operation in Eqn.(5). Given $z^{(n-1)}$ and $\beta^{(n-1)}$ which are outputs of previous layers in stage $n-1$, the output of this layer is defined as:

$$x^{(n)}=\left(A^TA+\rho^{(n)}I\right)^{-1}\left[A^Tb+\rho^{(n)}\left(z^{(n-1)}-\beta^{(n-1)}\right)\right] \tag{10}$$

where b is the input measurements and $\rho^{(n)}$ is learnable penalty parameter in nth stage. In the first stage, $z^{(0)}$ and $\beta^{(0)}$ are initialized to zeros, so the $x^{(1)}=\left(A^TA+\rho^{(1)}I\right)^{-1}\left(A^Tb\right)$.

Nonlinear transform layer ($\mathbf{Z^{(n)}}$): This layer performs nonlinear transform inspired by the shrinkage function $S(\cdot)$ defined in $\mathbf{Z^{(n)}}$ in Eqn.(5). Instead of setting it to be a fixed function determined by the regularization term, we learn a more general function by using a piecewise linear function. Given $x^{(n)}$ and $\beta^{(n-1)}$, the output of this layer is defined as:

$$z_l^{(n)}=S_{PLF}\left(x^{(n)}+\beta_l^{(n-1)};\left\{p_i,q_{l,i}^{(n)}\right\}_{i=1}^{N_c}\right) \tag{11}$$

where $S_{PLF}(\cdot)$ is a piecewise linear function determined by control points $\left\{p_i,q_{l,i}^{(n)}\right\}_{i=1}^{N_c}$.

$$S_{PLF}(a)=\begin{cases} a+q_{l,1}^{(n)}-p_1, & a<p_1 \\ a+q_{l,N_c}^{(n)}-p_{N_c}, & a>p_{N_c} \\ q_{l,r}^{(n)}+\frac{(a-p_r)\left(q_{l,r+1}^{(n)}-q_{l,r}^{(n)}\right)}{p_2-p_1}, & p_1\le a\le p_{N_c} \end{cases} \tag{12}$$

Multiplier update layer ($\mathbf{M^{(n)}}$): This layer is defined by the Lagrangian multiplier updating procedure $\mathbf{M^{(n)}}$ in Eqn.(5). The output of this layer in stage n is defined as:

$$\beta_l^{(n)}=\beta_l^{(n-1)}+\eta_l^{(n)}\left(x^{(n)}-z_l^{(n)}\right) \tag{13}$$

where η are learnable parameters.

D. Gradient Computation by Backpropagation

The parameters of ADMM-Net are optimized by the gradient-based algorithm L-BFGS [24]. In the forward pass,the data procession of n th stage is in the order of $\mathbf{X^{(n)}}$, $\mathbf{Z^{(n)}}$, $\mathbf{M^{(n)}}$. In the backward pass, the gradients are computed in an inverse order.

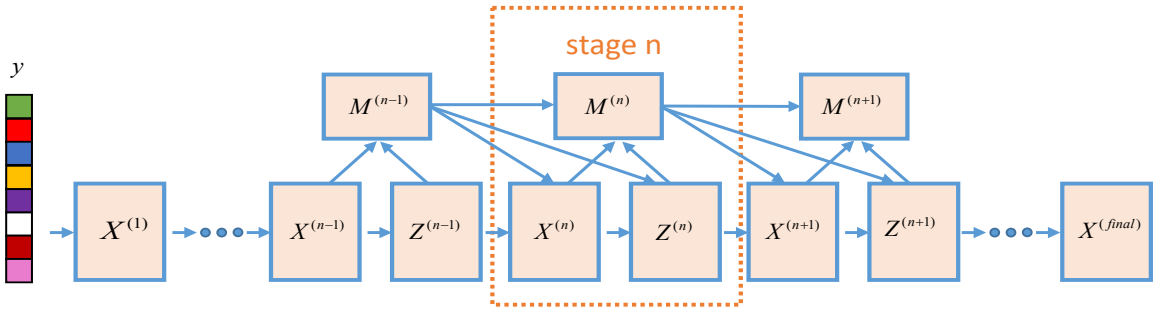

Fig. 1. Illustration of ADMM-Net. There are three types of nodes in n-th stage: reconstruction layer ($\mathbf{X}$), nonlinear transform layer ($\mathbf{Z}$) and multiplier update layer ($\mathbf{M}$)

.

III. Experiment

A. Experiment parameter set

We perform the Monte Carlo simulation in GATE to evaluate the effectiveness of proposed method. A hypothetical 430 mm diameter system with 32 modules, each module comprised 16×16 crystals array with 30 mm length and 2.4 mm pitch. The transaxial FOV is 330 mm, and axial FOV is 40 mm. The total crystal number is 8192 and material is LYSO. Since Defrise' work focus on the transaxial direction, thus in this paper, the number of detector unit is 512 rather than 8192. Temporal resolution is introduced into the system, which is a parameter of Gaussian blurring in the time domain. Background noise is also introduced, which the energy distribution is following a Gaussian distribution, and time distribution is following a Possion process. The simulated offsets were generated for the system, as the sum of two terms. The first term is generated as a uniform random number over the interval (-33.3, +33.3) ps as the floating offset for each crystal detector unit, and the second term is also generated as a uniform random number over interval (-33.3, +33.3) ps as the constant offset for each block array.

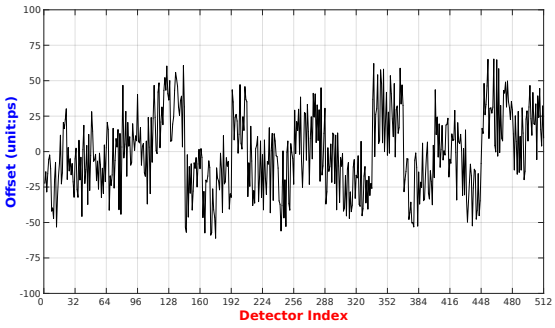

Fig. 2. The simulated offset x_j (unit: ps) for 512 detector unit

We perform three simulation experiments: point source experiment, cylinder source experiment, and hoffman phantom experiment. In point source experiment, 0.1 mm radius sphere point (1 Mbq ^{18}F) was placed at the center FOV, nearly one million coincidence events were used for calibration. For cylinder source experiment, a radius 0.5 mm, height 100 mm, line source (1 Mbq ^{18}F) is set. For the hoffman phantom experiment, the parameter is set as same as benchmark voxelized source in GATE. The scanning time in all experiments is 100s.

B. Result

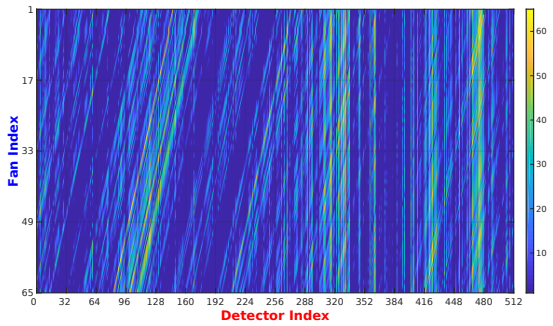

Fig. 3. The offsets calculated by method of consistency condition, the horizontal axial is detector index j, and vertical axis is fan index k

Table.I lists the parameters and results of experiments. In the point source, the effect of proposed method would be slightly inferior to traditional LSQR or Lasso, the FWHM is 328.54 ps and the FWHM of LSQR is 325.95 ps. While the distribution of source is more complex, the performance of proposed method is gradually emerging. In the situation of cylinder experiment, the index of FWHM of proposed method is nearly the traditional method, and in the phantom experiment, the performance of proposed method is better. Owing to that the label of network is based on the result of consistency condition method, the universality of method will be extended.

C. Other index

In this experiment, the label of network is using result of LSQR instead of result of consistency condition method. We found that the proposed method could obtain a relatively good result with slightly inferior comparing with result of traditional algorithm in long acquisition time. The result of experiment is list in Table.II. In some case, there is a need for some situation that the label of network is result of Lasso, then the computation time of proposed method will be

TABLE I

SIMULATION EXPERIMENT

Point source experiment				
Temporal resolution	Expect resolution	Noisy-data resolution	Noise(Poisson) (Exponetial)	Coin Window
200ps	323.12ps	346.07	2mus	2ns
Index \ Method	LSQR	Lasso	Consistency	ADMM-Net
FWHM(ps)	325.0709	325.9589	328.5403	327.9443
Cylinder source experiment				
Temporal resolution	Expect resolution	Noisy-data resolution	Noise(Poisson) (Exponetial)	Coin Window
200ps	946.2343ps	948.5801	2mus	2ns
Index \ Method	LSQR	Lasso	Consistency	ADMM-Net
FWHM(ps)	946.5363	946.5208	946.6387	946.58
Hoffman phantom experiment				
Temporal resolution	Expect resolution	Noisy-data resolution	Noise(Poisson) (Exponetial)	Coin Window
200ps	700.167ps	705.1116ps	2mus	2ns
Index \ Method	LSQR	Lasso	Consistency	ADMM-Net
FWHM(ps)	701.9131	701.8642	701.6629	701.6796

faster than traditional Lasso algorithm or other l_1 optimized method.

TABLE II

POINT EXPERIMENT WITH PARTIAL DATASET

Partial dataset for calibration				
data percent	100%	50%	33%	20%
Index \ Method	LSQR	LSQR	LSQR	LSQR
FWHM(ps)	325.0709	325.1301	326.2431	326.4732
data percent	10%	5%	1%	1%
Index \ Method	LSQR	LSQR	LSQR	ADMM-Net
FWHM(ps)	327.8715	328.6345	333.2029	325.0841

IV. COMPLIANCE WITH ETHICAL STANDARDS

The experimental procedures involving human subjects described in this paper were approved by the Institutional Review Board.

V. CONCLUSION

We proposed a end-to-end ADMM-Net deep network for timing calibration in PET. The method combines the advantage of consistency condition method and traditional ADMM optimization. In future work, we would explore the more details in the architecture of network in which will have impact on the performance of method.

ACKNOWLEDGMENT

This work was supported in part by the National Key Technology Research and Development Program of China (No: 2017YFE0104000, 2016YFC1300302). National Natural Science Foundation of China (U1809204, 61701436, 61525106).

REFERENCES

[1] C. J. Thompson, M. L. Camborde, and M. E. Casey. A central positron source to perform the timing alignment of detectors in a PET scanner. *IEEE Transactions on Nuclear Science*, 52(5):1300–1304, 2005.

[2] J. Hancock and C. J. Thompson. Evaluation of an instrument to improve PET timing alignment. *Nuclear Inst & Methods in Physics Research A*, 620(2-3):343–350, 2010.

[3] M. W. Lenox, T. Gremillion, S. Miller, and J. W. Young. Coincidence time alignment for planar pixellated positron emission tomography detector arrays. In *Nuclear Science Symposium Conference Record, 2001 IEEE*, 2001.

[4] D. Luo, J. J. Williams, M. K. Limkeman, M. J. Cook, and D. L. Mcdaniel. Crystal-based coincidence timing calibration for PET scanner. In *Nuclear Science Symposium Conference Record, 2002 IEEE*, 2002.

[5] A. E. Perkins, M. Werner, A. Kuhn, S. Surti, and J. S. Karp. Time of flight coincidence timing calibration techniques using radioactive sources. In *Nuclear Science Symposium Conference Record, 2005 IEEE*, 2005.

[6] X. Li. Timing calibration for Time-of-Flight PET using positron-emitting isotopes and annihilation targets. *IEEE Transactions on Nuclear Science*, 2016.

[7] M. Werner and J. Karp. TOF PET offset calibration from clinical data. *Physics in Medicine & Biology*, 58(12):4031–46, 2013.

[8] M. Defrise, A. Rezaei, and J. Nuyts. Time-of-Flight PET time calibration using data consistency. *Physics in Medicine & Biology*, 63(10), 2018.

[9] Y. Li. Consistency equations in native detector coordinates and timing calibration for time-of-flight PET. *Biomedical Physics & Engineering Express*, 5(2), 2018.

[10] S. J. Park, S. Southekal, M. Purschke, S. S. Junnarkar, and P. Vaska. Digital coincidence processing for the ratcap conscious rat brain PET scanner. *IEEE Transactions on Nuclear Science*, 55(1):510–515, 2008.

[11] A. B. Mann, Stephan Paul, Arne Tapfer, V. C. Spanoudaki, and Sibylle I. Ziegler. A computing efficient PET time calibration method based on pseudoinverse matrices. In *Nuclear Science Symposium Conference Record*, 2009.

[12] X. Yu, T Isobe, M Watanabe, and H. Liu. Sub-detector unit timing calibration for brain PET based on l1-norm constraint. *Biomedical Physics & Engineering Express*, 2(3):035008, 2016.

[13] X. Yu, T. Isobe, M. Watanabe, and H. Liu. Novel crystal timing calibration method based on total variation. *Physics in Medicine & Biology*, 61(22):7833, 2016.

[14] P. D. Reynolds. Convex optimization of coincidence time resolution for a high-resolution PET system. *IEEE Transactions on Medical Imaging*, 30(2):391–400, 2011.

[15] Christopher C. Paige and Michael A. Saunders. Lsqr: An algorithm for sparse linear equations and sparse least squares. *Acm Transactions on Mathematical Software*, 8(1):43–71, 1982.

[16] D. Fong and M. Saunders. LSMR: An iterative algorithm for sparse least-squares problems. *Siam Journal on Scientific Computing*, 33(5):2950–2971, 2010.

[17] D. L. Freese, Hsu. Dfc, D. Innes, and C. S. Levin. Robust timing calibration for PET using L1-norm minimization. *IEEE Transactions on Medical Imaging*, PP(99):1–1, 2017.

[18] H. Chen and H. Liu. PET timing calibration using low rank constraint. *Nuclear Instruments & Methods in Physics Research*, 987(5):164817, 2021.

[19] J. Zhang and B. Ghanem. ISTA-Net: Iterative shrinkage-thresholding algorithm inspired deep network for image compressive sensing. 2017.

[20] Y. Yang, J. Sun, H. Li, and Z. Xu. ADMM-CSNet: A deep learning approach for image compressive sensing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 521–538, 2020.

[21] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations & Trends® in Machine Learning*, 3(1):1–122, 2010.

[22] N. Parikh and S. Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, 2013.

[23] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58(1), 1996.

[24] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 1995.