Semi-Supervised Energy Spectrum Estimation in High Count-Rate Nuclear Spectrometry

Congyu Lin, Yiwei Huang, Zikang Chen, Songqi Gu,
Dima Bykhovsky, Tom Trigano, Xiaoying Zheng, Yongxin Zhu

***Abstract*—In nuclear spectroscopy, a physical phenomenon known as the pile-up effect distorts the direct measurements, which degrades the quality of energy and arrival time information. Since severely piled-up experimental signals cannot be reliably annotated, existing deep learning–based approaches rely predominantly on simulated training data. This leads to a limited generalization to real experimental measurement as a result of the simulation-to-real domain gap. To bridge this gap, we propose a semi-supervised learning framework that integrates experimental data into model training. It consists of three key components. First, simulated data and low count-rate experimental data are used to generate pseudo-labels for high count-rate experimental signals. Second, physics-guided constraints are applied to refine the pseudo-labels, ensuring physical consistency. Third, a stage-wise loss function is introduced to mitigate class imbalance and prevent model collapse. The proposed framework is experimentally validated on X-ray fluorescence (XRF) measurements acquired at the SSRF (Shanghai Synchrotron Radiation Facility). Comprehensive evaluations across a wide range of count-rate conditions demonstrate that the proposed method achieves superior performance in both spectrum estimation and count rate estimation compared to current state-of-the-art approaches.**

***Index Terms*—Nuclear spectroscopy, Semi-Supervised learning, Deep learning, Signal processing, Pile-up**

I. INTRODUCTION

Nuclear spectroscopy is a key experimental technique for investigating the structure and properties of atomic nuclei [1], [2]. During radiation detection, the inter-arrival times of photons can be shorter than the durations of the pulses generated during amplification, especially under high count-rate conditions. This temporal overlap, known as the *pile-up* effect, degrades the quality of energy and arrival time information, severely compromising measurement accuracy [3], [4].

This research was supported by the National Natural Science Foundation of China under grant numbers 12475196 and 12373113. (Corresponding authors: Dima Bykhovsky; Xiaoying Zheng)

Congyu Lin, Zikang Chen, Xiaoying Zheng, and Yongxin Zhu are with the Shanghai Advanced Research Institute, University of Chinese Academy of Sciences, Beijing 101408, China, and also with Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201210, China (e-mail: lincy@sari.ac.cn; chenzk@sari.ac.cn; zhengxy@sari.ac.cn; zhuyongxin@sari.ac.cn).

Yiwei Huang is with the Fudan University, Shanghai 200433, China(e-mail:25113060015@m.fudan.edu.cn)

Songqi Gu is with the Shanghai Synchrotron Radiation Facility, Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai, 201204, China(e-mail:gusq@sari.ac.cn)

Dima Bykhovsky is with the Department of Electrical and Electronics Engineering, Sami Shamoon College of Engineering, Be'er Sheva 8410802, Israel (e-mail: dmitrby@ac.sce.ac.il).

Tom Trigano is with the Department of Electrical and Electronics Engineering, Sami Shamoon College of Engineering, Ashdod 7765181, Israel (e-mail: thomast@ac.sce.ac.il).

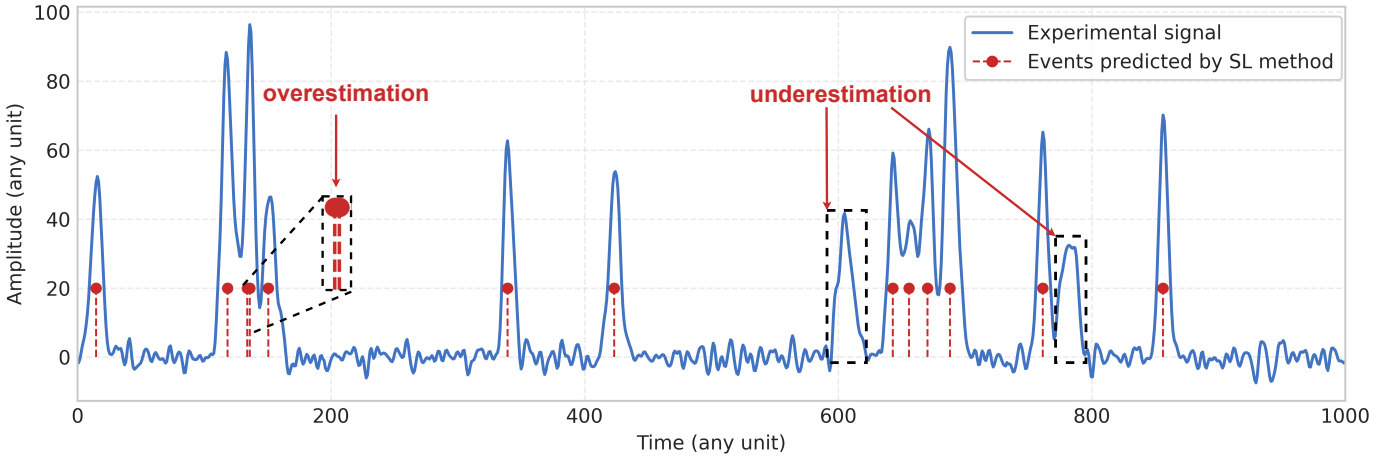
Extensive efforts have been devoted to pile-up correction [5]. Pile-up rejection approaches are commonly adopted to reduce misclassification of overlapping pulses, at the cost of discarding a portion of valid events [6], [7]. Other representative pile-up correction approaches include maximum likelihood estimation [8], [9], high-yield pile-up event recovery [10], [11], template matching [12], single-event reconstruction techniques [13], [14], and sparse regression-based modeling approaches [15], [16]. These approaches are generally effective but perform best under low count-rate conditions and relatively stable pulse shapes. More recently, statistical approaches have been introduced to mitigate these limitations by analyzing ensembles of pulses rather than detecting individual events in the time domain [17]–[20]. However, such methods typically require long data acquisition periods, which limits their applicability in real-time scenarios.

In the past decade, deep learning has emerged as a powerful tool in nuclear spectroscopy. Convolutional neural networks (CNNs) have demonstrated strong capability in pulse-shape feature extraction [21], [22]. Long short-term memory (LSTM) networks have been employed to capture temporal dependencies in sequential signals [23]. More recently, attention mechanisms have been introduced to further enhance feature representation and long-range dependency modeling [24]. In addition, several studies have explored hybrid approaches that combine physical modeling with deep learning frameworks to improve interpretability and robustness [25], [26].

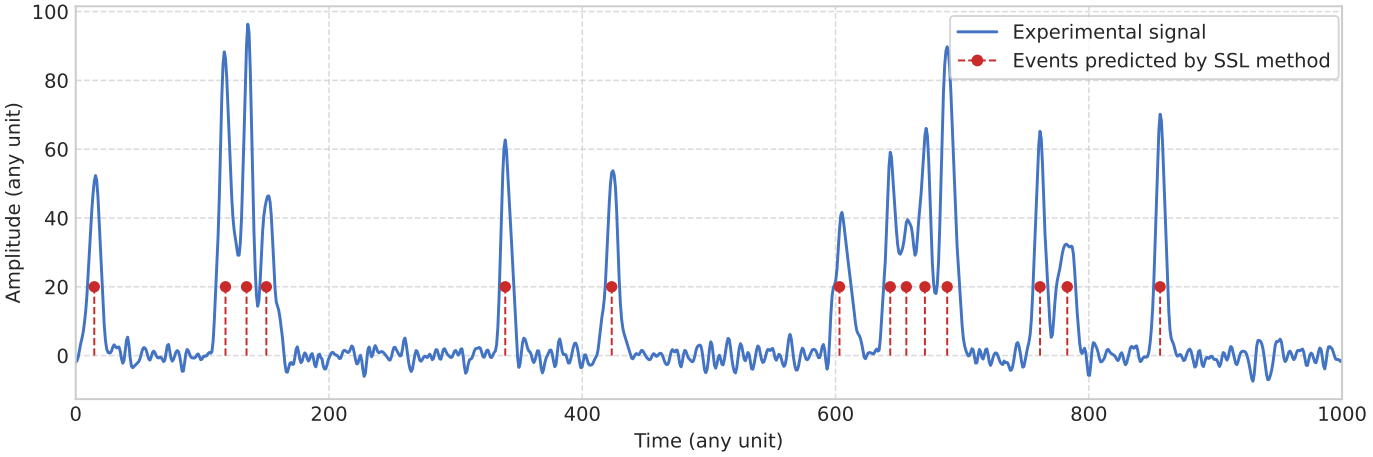
Under high count-rate conditions, severe pulse pile up prevents the acquisition of reliable ground-truth labels from experimental data. As a result, most current deep-learning-based methods are trained exclusively on simulated data. Nevertheless, simulations are inherently imperfect due to uncertainties in theoretical models, approximations in the underlying physical processes, and detector effects that are not fully captured [27], [28]. Consequently, simulated and experimental data often exhibit a distribution shift. As illustrated in Fig. 1a, models trained solely on simulated data lead to noticeable estimation errors when applied to real-world measurements.

In this paper, we incorporate high count-rate experimental data into deep learning training through a semi-supervised learning framework to mitigate these challenges. As illustrated in Fig. 1b, the proposed semi-supervised framework achieves superior performance on real measurements compared with its fully supervised counterpart. Specifically, we train a teacher model that leverages labeled low count-rate experimental data to learn intrinsic pulse shapes, while utilizing labeled high count-rate simulated data to capture pile-up patterns. The

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2026.3706156



(a)



(b)

Fig. 1: Comparison of event prediction results obtained using fully supervised and semi-supervised methods on real measurements. (a) Events predicted by fully-supervised method trained on simulated data. (b) Events predicted by the proposed semi-supervised method. Both methods employ AttnUNet++ [24] as the backbone network.

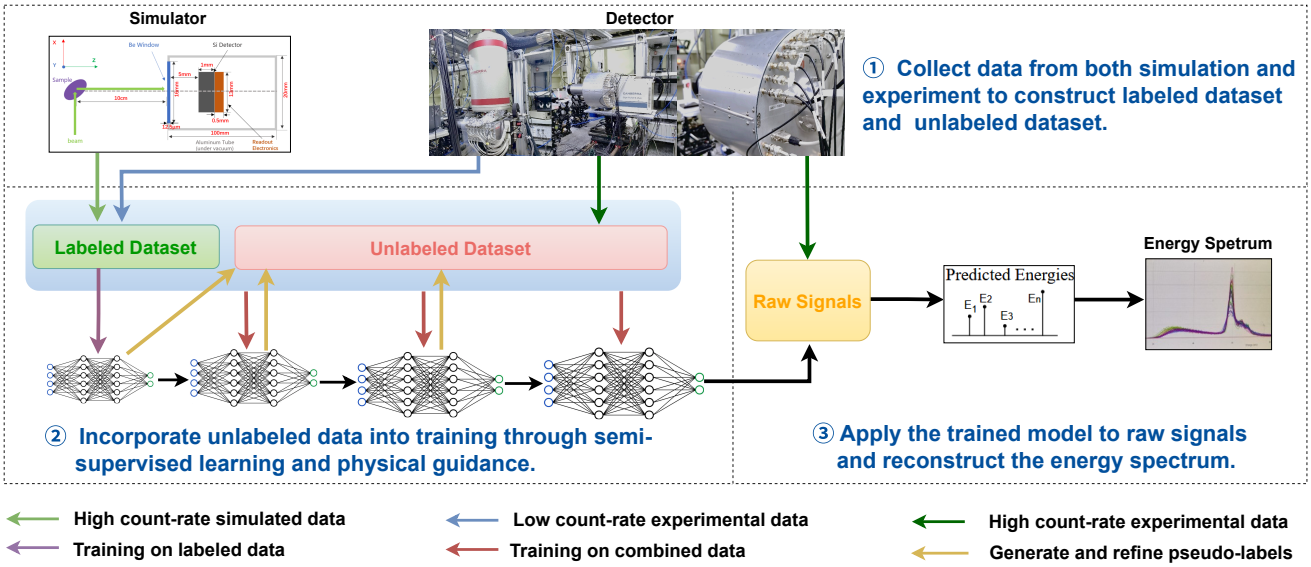


Fig. 2: Overview of the proposed semi-supervised spectrum estimation framework.

teacher model is then used to generate pseudo-labels for unlabeled high count-rate experimental signals. As illustrated in Fig. 2, the semi-supervised training proceeds iteratively as follows: (1) a teacher model is trained on the labeled dataset, which includes high count-rate simulated data and low count-rate experimental data; (2) the teacher generates pseudo-labels for unlabeled high count-rate signals; (3) the pseudo-labeled high count-rate data are combined with the original labeled dataset to train a student model; and (4) the student model is promoted to a new teacher, and the process is repeated. Through this iterative scheme, previously unlabeled high count-rate signals are progressively incorporated into training, enabling the model to learn more complex and diverse pulse characteristics.

Furthermore, nuclear radiation detection is inherently governed by physical laws, whereas purely data-driven pseudo-labeling may produce physically inconsistent labels. To address this issue, we propose a physics-guided pseudo-label filtering approach. Specifically, we first determine a physically plausible range of event counts by considering the pulse rise time and count rate. This criterion is then used to filter pseudo-labels, effectively removing physically inconsistent samples and ensuring that the final generated data remain consistent with the underlying physical processes.

Finally, the nuclear spectroscopy data exhibit intrinsic class imbalance. The background class (no event arrivals) vastly outnumbers event classes (valid event arrivals), and the latter are further skewed by the non-uniform photon energy distribution [23], [29]. Moreover, during semi-supervised training on severely imbalanced classes, there is an increased risk of model collapse [30]. To address these issues, we design a stage-wise loss function. Specifically, in the initial semi-supervised iteration, focal loss (FL) [31] is used to emphasize minority and hard-to-predict events, ensuring the model learns robust representations despite severe class imbalance. In subsequent semi-supervised iterations, the cross-entropy (CE) loss is combined with a KL-divergence [32] regularization term. This KL constraint encourages the student model to mimic the teacher's label distribution, preventing it from collapsing into overconfident predictions.

The main contributions of this work are as follows:

- We propose a semi-supervised framework that enhances model performance and generalization on real measurements by incorporating unlabeled, high count-rate experimental data into the training process.
- A novel physics-guided refinement strategy is introduced, which uses Poisson arrival statistics and detector physics to filter and reject physically unreasonable pseudo-labels.
- A stage-wise loss function is designed, using Focal Loss to combat initial class imbalance and KL-Divergence regularization to prevent model collapse during student training.
- We validate the effectiveness of the proposed framework through extensive experiments conducted at SSRF (Shanghai Synchrotron Radiation Facility). Comprehensive evaluations demonstrate that our method consistently outperforms state-of-the-art approaches in both spectrum estimation and count rate estimation.

TABLE I: Summary of the notations used in this work.

Variable	Definition
$\mathcal{X}$	Raw signal dataset (Unlabeled), i.e., $\{X_i\}_{i=1}^{n}$
$\mathcal{Y}$	Pseudo-labels for $\mathcal{X}$, i.e., $\{Y_i\}_{i=1}^{n}$
$(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}})$	Labeled dataset, i.e., $\{(\tilde{X}_i, \tilde{Y}_i)\}_{i=1}^{\tilde{n}}$
$(\mathcal{X}', \mathcal{Y}')$	Refined pseudo-labeled dataset, i.e., $\{(X_i', Y_i')\}_{i=1}^{n'}$
λ	Count rate (cps)
ρ	Normalized intensity
γ	Focal loss focusing parameter
α	KL-divergence regularization coefficient
N	Number of photons; $N(t)$ denotes the number of photons within time t
$\hat{N}$	Estimated number of events; $\hat{N}(x)$ denotes the number of photons generating peak x
$N_{\max}$	Maximum physically plausible photon count
M	Length of the signal; $M = 1024$ in this work
θ	Model parameters; θ_t and θ_s denote teacher and student models, respectively
θ^*	Optimal model parameters; θ_t^* and θ_s^* denote teacher and student optima
$\mathcal{S}(\cdot)$	Pseudo-label refinement function (Algorithm 3)
$\mathcal{L}(\cdot)$	Loss function; includes focal loss $\mathcal{L}_{FL}$ and cross-entropy loss $\mathcal{L}_{CE}$
$\mathcal{R}(\cdot)$	Regularization term; $\mathcal{R}_{KL}(\cdot)$ denote Kullback–Leibler divergence
$\mathrm{PA}(\cdot)$	Pulse area analysis (Section II-A)
$\mathcal{K}_i$	Set of detected peaks in $\tilde{X}_i$
T_{thresh}	Amplitude threshold

The main impact of this work is a novel semi-supervised framework that successfully integrates unlabeled, high-rate experimental data, enabling the model to learn and recognize complex, real-world pulse shapes that fully-supervised methods, trained only on simulated data, fail to identify.

II. Problem Formulation

In this section, we first present the principle of pile-up from a stochastic-process perspective. We then formally define the problem of estimating the energy spectrum from the perspective of deep learning. Table I summarizes some of the frequently used notations throughout this paper.

A. Signal Modeling

The detector output can be theoretically modeled as a linear superposition of pulse shapes [19]. Formally, the response of detector is expressed as

$$r(t) = \sum_{j=-\infty}^{\infty} a_j \, \delta(t - \tau_j) * \Phi_j(t) + \varepsilon(t), \tag{1}$$

where the photon arrival times, τ_j, are unknown and are assumed to follow a Poisson process with count rate parameter λ [33]. Each photon arrival at time τ_j is modeled as a Dirac delta function with amplitude a_j, and $\varepsilon(t)$ represents additive noise. $\Phi_j(t)$ is the pulse shape, which depends on the type of

detector and pulse shaper. In practical digital pulse processing systems, the continuous signal $r(t)$ is sampled by an analog-to-digital converter (ADC) at discrete time step $t_i = i\Delta t$, yielding a discrete sequence $v(i) = r(t_i)$.

In the absence of pile-up, the photon energy can be estimated using pulse area analysis (PAA) [34]. Specifically, the pulse area is computed as

$$PA \approx \sum_{i=0}^{m} v_0(i), \tag{2}$$

where $v_0(i)$ denotes the baseline-corrected sampled signal and m is the number of samples within the pulse window. The pulse area is proportional to the total collected charge and, consequently, to the energy deposited in the detector.

However, under high count-rate conditions, pile-up distorts the mapping between pulse area and photon energy. Accurate estimation of photon energy at high count-rates, therefore, becomes the primary objective of this work.

B. Spectrum Estimation Modeling

From a deep learning perspective, a sequence of raw signals generated by the detector can be represented as a one-dimensional time series X_i of length M:

$$X_i = \{x_1, x_2, x_3, \ldots, x_M\}. \tag{3}$$

The corresponding ground-truth label Y_i^{true} for X_i is defined as a vector of the same length, encoding the energy of each detected photon:

$$Y_i^{\text{true}} = \{y_1, y_2, y_3, \ldots, y_M\}, \quad y_l = \begin{cases} \mathcal{E}_k, & l = \tau_k,\ k \in [1, N], \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Here, τ_k denotes the k-th photon arrival time, and $\mathcal{E}_k$ represents its corresponding energy.

Consequently, the task of spectrum estimation can be formulated as a classification problem, in which the objective is to predict the photon energy y_l at each time step x_l, or assign $y_l = 0$ when no photon is detected. Under a fully supervised setting, this problem reduces to empirical risk minimization over labeled data. The optimal parameters θ^* are obtained by:

$$\theta^* = \arg\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}\big(Y_i^{\text{true}}, f(X_i, \theta)\big), \tag{5}$$

where $\mathcal{L}(\cdot)$ denotes the loss function.

However, as discussed in Section II-A, only simulation data and low count-rate experimental data can be reliably labeled. These samples constitute a labeled dataset, denoted as $(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}) = \{(\tilde{X}_i, \tilde{Y}_i)\}_{i=1}^{\tilde{n}}$, where $\tilde{X}_i$ and $\tilde{Y}_i$ represent the i-th signal and its corresponding label, respectively, and $\tilde{n}$ is the number of labeled samples.

In contrast, ground-truth labels for high count-rate experimental data are generally unavailable. We denote this unlabeled dataset as $\mathcal{X} = \{X_i\}_{i=1}^{n}$, where X_i is the i-th signal and n is the number of unlabeled samples. This limitation motivates a transition from fully supervised learning to a semi-supervised framework, where both the labeled dataset $(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}})$ and the unlabeled dataset $\mathcal{X}$ are utilized during training.

From a modeling perspective, the challenge lies in the fact that the empirical risk over $\mathcal{X}$ cannot be directly evaluated due to the absence of labels. This motivates the introduction of an auxiliary labeling mechanism that generates pseudo-labels $\mathcal{Y}$ for $\mathcal{X}$, enabling risk minimization to be extended to the unlabeled domain.

Since such pseudo-labels are inherently noisy, it is further necessary to consider a refinement operator $\mathcal{S}(\cdot)$ that maps the initial pseudo-labeled dataset to a more reliable subset:

$$(\mathcal{X}', \mathcal{Y}') = \mathcal{S}\big(\mathcal{X}, \mathcal{Y}\big), \tag{6}$$

where $(\mathcal{X}', \mathcal{Y}') = \{(X_i', Y_i')\}_{i=1}^{n'}$ denotes the refined pseudo-labeled dataset.

Under this formulation, the learning objective can be generalized as a joint empirical risk over labeled and pseudo-labeled data:

$$\begin{aligned} \theta^* = \arg\min_{\theta} \Bigg[& \frac{1}{\tilde{n}} \sum_{i=1}^{\tilde{n}} \mathcal{L}\big(\tilde{Y}_i, f(\tilde{X}_i, \theta)\big) \\ & + \frac{1}{n'} \sum_{i=1}^{n'} \mathcal{L}\big(Y_i', f(X_i', \theta)\big) \\ & + \alpha \mathcal{R}(\mathcal{X}', \theta) \Bigg], \end{aligned} \tag{7}$$

where $\mathcal{R}(\cdot)$ is the regularization term, and α controls its contribution.

III. Methodology

In this section, we describe the technical details of the proposed semi-supervised learning framework. The overall workflow is first illustrated in Algorithm 1 and Section III-A. Then, the pseudo-label filtering strategy $\mathcal{S}(\cdot)$ is detailed in Section III-B, and summarized in Algorithm 3. The loss function $\mathcal{L}(\cdot)$ and the regularization term $\mathcal{R}(\cdot)$ are described in Section III-C. Finally, the backbone network selection is presented in Section III-D.

A. Semi-Supervised Learning Framework

As illustrated in Fig. 3, the semi-supervised training process begins by training a teacher model on a labeled dataset $(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}})$, which consists of both simulated data and low count-rate experimental data. Formally,

$$(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}) = (\mathcal{X}^{\text{sim}}, \mathcal{Y}^{\text{sim}}) \cup (\mathcal{X}^{\text{exp\_l}}, \mathcal{Y}^{\text{exp\_l}}), \tag{8}$$

The motivation for the dataset composition is that the teacher can simultaneously learn realistic pulse shapes from experimental data and representative pile-up patterns from simulated data.

It is noteworthy that pile-up may still occur, although not severe, at low count-rates due to the stochastic nature of photon arrivals. To obtain reliable energy labels under such conditions, a low count-rate experimental data labeling strategy is designed. Specifically, each recorded signal is first divided into smaller segments according to its zero-crossing points and a predefined amplitude threshold. Signals that contain multiple peaks within any segment are then excluded

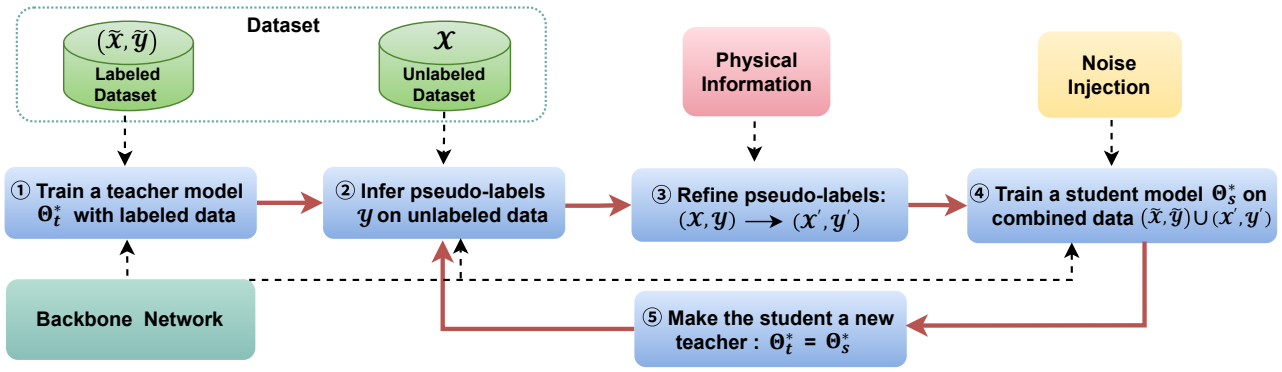


Fig. 3: Semi-supervised training pipeline. A teacher model trained on labeled data produces pseudo labels for unlabeled samples. These pseudo labels are refined using physics-guided constraints, combined with labeled data to train a student model. The student model is then iteratively updated as the new teacher.

Algorithm 1 Semi-Supervised Learning for Nuclear Spectrum Estimation

Require: Labeled dataset $(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}})$ and unlabeled dataset $\mathcal{X}$.
Ensure: optimal student model parameter θ_s^*

1: Learn teacher model θ_t^* by minimizing loss on labeled dataset:

$$\theta_t^* = \arg\min_{\theta_t} \frac{1}{\tilde{n}} \sum_{i=1}^{\tilde{n}} \mathcal{L}(\tilde{Y}_i, f(\tilde{X}_i, \theta_t))$$

2: Generate pseudo labels for unlabeled dataset:

$$Y_i = f(X_i, \theta_t^*)$$

3: Refine the pseudo labels according to physical information:

$$(X_i', Y_i') = \mathcal{S}(X_i, Y_i)$$

4: Learn a student model θ_*^s by minimizing the loss on labeled data and filtered pseudo-labeled data, with a regularization term:

$$\theta_s^* = \arg\min_{\theta_s} \Bigg[\frac{1}{\tilde{n}} \sum_{i=1}^{\tilde{n}} \mathcal{L}\big(\tilde{Y}_i, f(\tilde{X}_i, \theta_s)\big) + \frac{1}{n'} \sum_{i=1}^{n'} \mathcal{L}\big(Y_i', f(X_i', \theta_s)\big) + \alpha \mathcal{R}\Big(\mathcal{X}', \theta_s\Big) \Bigg]$$

5: Use the student as the new teacher and return to Step 2 until convergence.

to eliminate potential pile-up artifacts. For the remaining clean signals, the corresponding energy labels are computed using pulse-area analysis. The detailed steps of this procedure are summarized in Algorithm 2.

Once the teacher model is trained, it is used to generate pseudo-labels $\mathcal{Y}$ for high count-rate experimental data by performing inference on the unlabeled dataset $\mathcal{X}$. Specifically, each time point is pseudo-labeled as the class with the maximum predicted probability among all categories.

Afterward, the refined pseudo labeled data $(\mathcal{X}', \mathcal{Y}')$ are obtained based on physical constraints, as detailed in Section III-B. These refined data are then merged with the original

Algorithm 2 Low Count-Rate Experimental Data Labeling Strategy

Require: Low count-rate experimental signals $\mathcal{X}^{\text{exp\_l}}$
Ensure: Labeled low count-rate experimental dataset $(\mathcal{X}^{\text{exp\_l}}, \mathcal{Y}^{\text{exp\_l}})$

1: **for** each signal $X_i^{\text{exp\_l}} \in \mathcal{X}^{\text{exp\_l}}$ **do**
2: Segment $X_i^{\text{exp\_l}}$ into a set of sub-signals $\{s_{i,1}, s_{i,2}, \ldots, s_{i,K_i}\}$ based on zero-crossing points and an amplitude threshold:

$$X_i^{\text{exp\_l}} \xrightarrow[\text{zero-crossings}]{\text{segmentation}} \{\, s_{i,j} \mid \max(s_{i,j}) > T_{thresh} \,\}.$$

3: Exclude $X_i^{\text{exp\_l}}$ if any segment $s_{i,j}$ contains multiple peaks:
4: Compute the energy label of the remaining signals via pulse-area analysis:

$$Y_i^{\text{exp\_l}} = PA(X_i^{\text{exp\_l}}).$$

5: **end for**
6: **return** Labeled dataset $(\mathcal{X}^{\text{exp\_l}}, \mathcal{Y}^{\text{exp\_l}})$

labeled dataset to construct a combined dataset:

$$(\mathcal{X}_c, \mathcal{Y}_c) = (\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}) \cup (\mathcal{X}', \mathcal{Y}'). \tag{9}$$

Subsequently, a student model with equal or larger scale is trained on the combined dataset. During student training, physically motivated noise injection is applied as a data augmentation strategy to enhance model robustness. Specifically, additive white noise is introduced to emulate electronic noise [35], and a Gaussian filtering with an L_1-normalized Gaussian kernel is performed to mimic the detector's bandwidth limitation [36]. These two operations not only diversify the pulse shapes but also preserve the pulse areas, thus maintaining label consistency before and after augmentation.

The trained student is then promoted to a new teacher to generate updated pseudo-labels for $\mathcal{X}$. This procedure is repeated iteratively, during which the model is expected to progressively improve its feature extraction capability and generalization to high count-rate experimental data.

B. Physics-Guided Pseudo-label Refinement Strategy

In the classic teacher-student paradigm of semi-supervised learning, the model is responsible for generating pseudo-

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2026.3706156

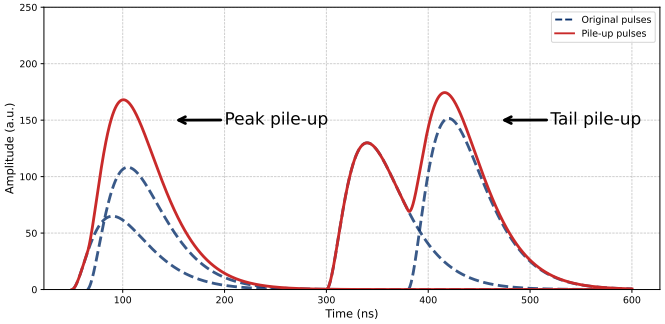


Fig. 4: Illustration of peak pile-up and tail pile-up. In tail pile-up, each pulse peak corresponds to one photon event, whereas in peak pile-up, a single peak originates from multiple photon events.

labels for real experimental data. However, this purely data-driven process relies solely on parameters learned by the network and lacks consideration of physical principles. This can easily lead to labels that are inconsistent with realistic physical meaning. To address this, we introduce a pseudo-label filtering mechanism based on stochastic process theory to reject unreasonable pseudo-labels, thereby effectively incorporating physical information within the semi-supervised learning process.

Before presenting the proposed strategy, we first introduce two types of pile-up phenomena and the corresponding probabilistic models, as illustrated in Fig. 4. **Peak pile-up** arises when the photon inter-arrival time is shorter than the pulse rise time, causing multiple photons to merge into a single local maximum (peak) in the resulting waveform. In contrast, **Tail pile-up** occurs when the photon inter-arrival time exceeds the pulse rise time but remains shorter than the total pulse duration, resulting in distinct local maxima(peaks) for each photon in the combined waveform. Therefore, let $\hat{N}(x)$ donate the number of photons that contribute to the peak x. It satisfies the following requirement.

$$1 \leq \hat{N}(x) \leq N_{\max}, \tag{10}$$

where $\hat{N}(x) = 1$ corresponds to either no pile-up or tail pile-up, whereas $1 < \hat{N}(x) \leq N_{\max}$ indicates peak pile-up. $N_{\max}$ is the maximum physically plausible photon count for a peak pile-up. Our overall objective is thus to estimate $N_{\max}$ and $\hat{N}(x)$, and then filter pseudo-labels according to this constraint.

Algorithm 3 summarizes the pseudo-label filtering process. First, $N_{\max}$ is defined as the smallest integer such that the probability of observing more than $N_{\max}$ photons contributing to a peak is less than 10^{-4}.

Assuming photon arrivals follow a Poisson process with count rate λ, the number of photons within the rise time t_r, conditioned on at least one arrival, follows a zero-truncated Poisson distribution:

$$\mathbb{P}(N(t_r) = k \mid N(t_r) \geq 1) = \frac{(\lambda t_r)^k e^{-\lambda t_r}}{k!\,(1 - e^{-\lambda t_r})}. \tag{11}$$

Accordingly, $N_{\max}$ is determined as:

$$N_{\max} = \min \left\{ k \in \mathbb{Z}^+ : \mathbb{P}\big(N(t_r) > k \mid N(t_r) \geq 1\big) < 10^{-4} \right\}. \tag{12}$$

Next, for each pseudo-labeled data pair (X_i, Y_i), the algorithm identifies a set of candidate peaks, denoted as $\mathcal{K}_i$, which correspond to local maxima exceeding the amplitude threshold T_{thresh}.

For each peak $x_{i,k} \in \mathcal{K}_i$, we estimate the corresponding photon count $\hat{N}(x_{i,k})$, defined as the number of non-zero pseudo-labels within a temporal window of length $2t_r$ centered at the peak. Specifically,

$$\hat{N}(x_{i,k}) = \big|\{\, y_{i,t} > 0 \;\big|\; t \in [t_k - t_r,\ t_k + t_r]\,\}\big|, \tag{13}$$

where $y_{i,t}$ denotes the t-th entry of the pseudo-label sequence Y_i, and t_k is the time index of the peak $x_{i,k}$.

Finally, a pseudo-labeled data pair (X_i, Y_i) is retained in the refined dataset if all detected peaks satisfy the following constraint:

$$1 \leq \hat{N}(x_{i,k}) \leq N_{\max}, \quad \forall x_{i,k} \in \mathcal{K}_i. \tag{14}$$

For better understanding, we provide an example. Consider an SDD with an average rise time of $t_{\text{r}} = 80$ ns, operating under experimental conditions where the input count rate is lower than 5 Mcps, corresponding to a count rate of $\lambda = 5 \times 10^6$ s^{-1}. Under these conditions,

$$\begin{aligned} &\mathbb{P}\big(N(t_r) > 4 \mid N(t_r) \geq 1\big) > 10^{-4}, \\ &\mathbb{P}\big(N(t_r) > 5 \mid N(t_r) \geq 1\big) < 10^{-4}. \end{aligned} \tag{15}$$

This yields a maximum physically plausible photon count of $N_{\max} = 5$.

Next, we detect peaks $x_{i,k}$ in X_i and estimate $\hat{N}(x_{i,k})$ by counting the number of nonzero photon labels within a temporal window $[t_k - t_r,\ t_k + t_r]$ of Y_i. Accordingly, a data pair (X_i, Y_i) is retained only if

$$1 \leq \hat{N}(x_{i,k}) \leq 5, \quad \forall\, x_{i,k} \in \mathcal{K}_i.$$

Since t_{r} is detector-specific and λ varies with experimental conditions, these two parameters are set as user-defined variables. Users can specify their values according to the detector type and experimental setup, after which the algorithm automatically determines an appropriate pseudo-label threshold. Through this adaptive threshold mechanism, the proposed method achieves broad applicability across diverse detection systems and experimental configurations.

C. Stage-wise Loss Function

The nuclear spectroscopic dataset exhibits a class imbalance arising from the physical properties of the recorded signals [23], [24]. As indicated by the signal characteristics in Eq. (1) and the labeling strategy in Eq. (4), only a small fraction of each recorded signal contains photon arrivals, causing the background class (i.e., class 0) to dominate the dataset. Moreover, within the non-zero classes, the distribution remains highly imbalanced due to the non-uniform photon energy distribution. Taking the XRF measurement of Cu as an

Algorithm 3 Physics-Guided Pseudo-Label Filtering

Require: Dataset $(\mathcal{X}, \mathcal{Y})$, count rate λ, pulse rise time t_r, amplitude threshold T_{thresh}.
Ensure: Refined dataset $(\mathcal{X}', \mathcal{Y}')$.
1: Compute maximum physically plausible photon count $N_{\max}$ using Eq. (11) and Eq. (12).
2: **for** each signal $X_i \in \mathcal{X}$ **do**
3: Identify candidate peaks: $\mathcal{K}_i = \{x_{i,t} \mid x_{i,t}$ is a local maximum and $x_{i,t} > T_{\text{thresh}}\}$
4: Estimate photon count $\hat{N}(x_{i,k})$ for each peak using Eq. (13).
5: Retain sample (X_i, Y_i) if constraint in Eq. (14) is satisfied.
6: **end for**
7: **return** $(\tilde{\mathcal{X}}', \tilde{\mathcal{Y}}')$

example, a large number of samples are concentrated around the $K_{\alpha 1}$ and $K_{\beta 1}$ emission lines, while significantly fewer samples occur in other energy regions. This imbalance poses a challenge for accurately classifying sparsely populated energy regions.

In addition to data imbalance, semi-supervised learning introduces another potential issue known as model collapse [30]. It is typically triggered when pseudo-labels generated from unlabeled data are highly uncertain or biased, causing the model to reinforce its own incorrect predictions in subsequent iterations. As a result, the learned feature representations become dominated by one or a few classes. This collapse poses a great challenge to maintain a stable energy distribution, as will be evident in Section IV-H.

To address these two challenges, we propose a stage-wise loss design. In the first semi-supervised iteration, the teacher model is trained on labeled data using Focal Loss (FL), which is defined as

$$\mathrm{FL} = -\sum_{i=1}^{C} y_i \, (1 - p_i)^{\gamma} \log(p_i), \tag{16}$$

where C is the number of classes, y_i is the one-hot encoded ground-truth label, p_i denotes the predicted probability for class i, and $\gamma \geq 0$ is the focusing parameter When $\gamma = 0$, the loss reduces to the standard cross-entropy loss.

By introducing the modulating term $(1 - p_i)^{\gamma}$, the loss dynamically reduces the contribution of easily classified samples, while assigning higher weights to misclassified or ambiguous ones. This mechanism forces the model to focus on harder, less confident examples that are often underrepresented in the dataset.

In the subsequent semi-supervised iterations, a Kullback–Leibler (KL) divergence regularization term is added to the cross-entropy (CE) loss to enforce consistency between the student and teacher predictions. Formally, the KL regularization term is defined as:

$$\mathcal{R}_{\mathrm{KL}} = \sum_{i=1}^{C} P_T(i) \log \frac{P_T(i)}{P_S(i)}, \tag{17}$$

where P_T and P_S denote the output probability distributions of the teacher and student models, respectively.

Overall, in the first iteration of the semi-supervised learning process, a teacher model is trained to minimize the focal loss on the labeled dataset:

$$\theta_t^* = \arg\min_{\theta_t} \frac{1}{\tilde{n}} \sum_{i=1}^{\tilde{n}} \mathcal{L}_{FL}\big(\tilde{Y}_i, f(\tilde{X}_i, \theta_t)\big). \tag{18}$$

In subsequent iterations, a student model is trained to minimize an loss function that combines three terms:

$$\begin{aligned} \theta_s^* = \arg\min_{\theta_s} \Bigg[& \frac{1}{\tilde{n}} \sum_{i=1}^{\tilde{n}} \mathcal{L}_{CE}\big(\tilde{Y}_i, f(\tilde{X}_i; \theta_s)\big) \\ & + \frac{1}{n'} \sum_{i=1}^{n'} \mathcal{L}_{CE}\big(Y_i', f(X_i'; \theta_s)\big) \\ & + \alpha \mathcal{R}_{KL}\big(f(\mathcal{X}', \theta_s)\big) \Bigg]. \end{aligned} \tag{19}$$

Here, the first term ensures learning on labeled data, the second term transfers knowledge from pseudo-labeled samples, and the third term stabilizes training by constraining the student's predictions to remain consistent with the teacher's output distribution. α is a regularization coefficient that controls the strength of the consistency constraint with the motivation to prevent class collapse and to reduce overconfidence in incorrect predictions.

D. Backbone Network

This semi-supervised learning framework can flexibly accommodate different backbone networks. As feature extractors, different backbones have a significant impact on both model performance and computational cost. We evaluate the effect of different backbones on count rate estimation and spectrum estimation in Section IV-C and Section IV-D, respectively. In addition, the computational complexity and inference latency of different backbones are analyzed in Section IV-I. In practical applications, the choice of backbone should be made by jointly considering accuracy requirements and resource constraints, such as computation cost, memory usage, and latency.

In Section IV-E, Section IV-F, Section IV-G, and Section IV-H, AttnUNet++ [24] is used as the default backbone unless otherwise specified.

IV. EXPERIMENTS AND EVALUATION

In this section, we first describe the experimental setup and the evaluation metrics. Subsequently, we compare our method with analytical and deep learning methods in terms of count rate estimation and spectrum estimation. This is followed by a quantitative analysis of the simulation-to-real gap. We then present a challenging case analysis, along with evaluations of the convergence behavior and an ablation study. Finally, we analyze the computational efficiency and practical applicability of the proposed method.

TABLE II: Experimental setup and configurations.

Setting	Value
Semi-supervised iterations	5
Labeled training samples(simulation)	5000
Labeled training samples(experiment)	5000
Unlabeled training samples	20000
Testing samples	5000
γ (Focal Loss focusing parameter)	1
α (KL regularization coefficient)	0.1
$N_{\max}$ (Maximum plausible photon count)	3

A. Experimental Setup

As shown in Fig. 5a, the experimental data were acquired from XRF measurements conducted at the SSRF, using a HITACHI Vortex-ME4 SDD equipped with a pulsed-reset preamplifier. The signals were digitized by a 250 MHz ADC and processed using a digital pulse processing (DPP) system. The digitized signals were low-pass filtered and subsequently downsampled by a factor of 2 to reduce data throughput, resulting in an effective temporal resolution of 8 ns per time step. Data were collected from copper (Cu) and cobalt (Co) under count rates of approximately 50 kcps, 80 kcps, 450 kcps, and 1 Mcps, respectively.

As illustrated in Fig. 5b, simulated data were generated using the Allpix Squared framework [37], which has been widely adopted in international high-energy physics experiments such as ATLAS [38] and CMS [39]. In this study, the simulated detector was configured to replicate the specifications of the HITACHI Vortex-ME4 SDD. Based on this configuration, XRF simulations were performed for Cu and Co samples under input count-rates consistent with the experimental conditions.

The labeled dataset comprises 5,000 experimental data at a count rate of 50 kcps processed using Algorithm 2, and 5,000 simulated data at a count rate of 1 Mcps. The unlabeled dataset consists of 20,000 experimental data with count rates of 80 kcps, 450 kcps, and 1 Mcps. For evaluation, additional 5,000 signals were selected as the testing dataset.

The model were trained on a single NVIDIA A30 GPU with 32 GB memory, using Python 3.10, PyTorch 2.3, and CUDA 12.1. The experimental settings are summarized in Table II, while a detailed discussion of parameter selection is provided in Section IV-J.

B. Performance Metrics

To comprehensively evaluate the proposed semi-supervised framework, we employ two complementary metrics: ICR and OCR for count rate estimation performance, and FWHM for spectrum reconstruction quality.

a) ICR and OCR: The input count rate (ICR) represents the true photon arrival rate incident on the detector, while the output count rate (OCR) denotes the number of pulses successfully identified after electronic processing. Due to pulse pile-up and system dead time, the ratio OCR/ICR is typically less than 1. Ideally, an OCR/ICR value close to 1 indicates accurate event recovery with minimal pulse loss. In the controlled experimental setting considered in this work, the ICR is known a priori, and thus a higher OCR directly reflects improved performance in count-rate estimation.

b) FWHM: The full width at half maximum (FWHM) of the reconstructed energy spectrum is used to quantify the system's energy resolution. A smaller FWHM indicates reduced spectral broadening and better compensation of pile-up-induced distortion. In this work, FWHM is computed from a smoothed spectrum obtained via kernel density estimation (KDE).

C. Count Rate Estimation

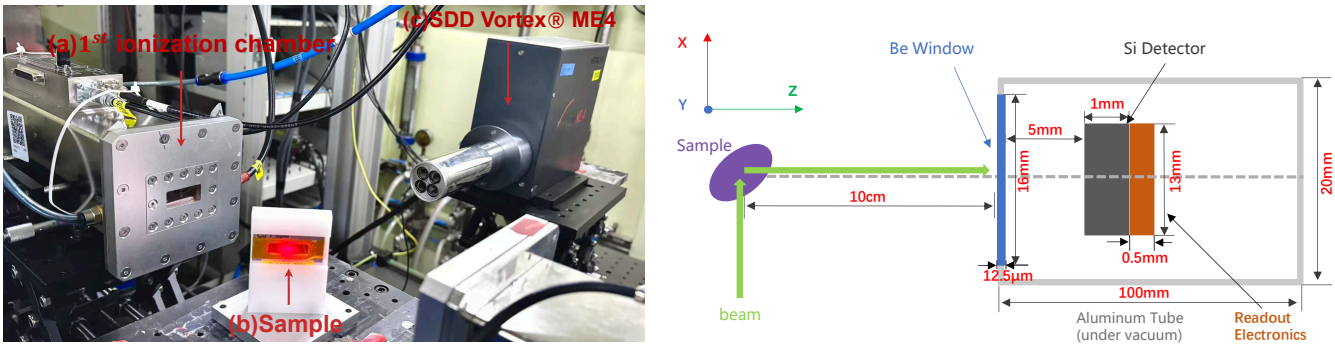
In this part, we compare the count rate estimation performance of three baseline methods: a classical analytical method [15] (denoted as Iterative LASSO), two deep learning-based method [24] (denoted as AttnUNet++), and NoA-DNN [25] (count rate estimation–specific method). To further evaluate the effectiveness of the proposed framework, we construct two semi-supervised variants by integrating it into AttnUNet++ and NoA-DNN, denoted as SSL-AttnUNet++ and SSL-NoA-DNN, respectively.

Experiments are conducted under multiple input count rate (ICR) conditions, including 88 kcps, 458 kcps, and 1142 kcps for Cu, and 81 kcps, 459 kcps, and 1115 kcps for Co. The output count rates (OCRs) of all five methods are evaluated under these conditions.

As shown in Fig. 6, in the low ICR regime (ICR $<$ 100 kcps), all methods achieve good counting performance, with OCR closely matching the true ICR. As the ICR increases, the deviation between OCR and ICR becomes more pronounced, reflecting the increasing difficulty in resolving overlapping pulses and complex temporal structures.

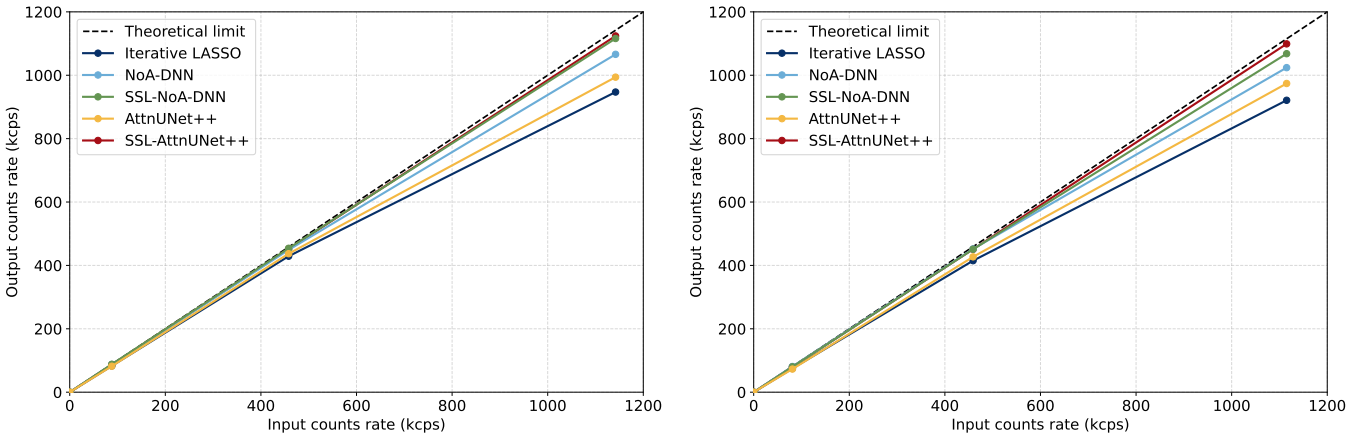
Detailed quantitative results are provided in Table III. Overall, deep learning-based methods consistently outperform the classical Iterative LASSO approach across all backbones. More importantly, across different backbone networks (AttnUNet++ and NoA-DNN), the proposed semi-supervised framework consistently improves count rate estimation performance compared with their fully supervised counterparts. Specifically, under high count rate conditions (ICR $>$ 1 Mcps), when using AttnUNet++ as the backbone, the proposed method improves OCR/ICR about 11%. When using NoA-DNN as the backbone, the improvement is about 4%. These results demonstrate that the proposed framework achieves superior count rate estimation accuracy and enhanced robustness under severe pile-up conditions.

In addition, under low count-rate conditions (ICR $<$ 1 Mcps), SSL-NoA-DNN achieves better performance than SSL-AttnUNet++, whereas the opposite trend is observed under high count-rate conditions. We hypothesize that this behavior arises from the different inductive biases of the two models. Although NoA-DNN has a relatively simpler architecture than AttnUNet++, it explicitly incorporates prior



(a) (b)

Fig. 5: Illustration of simulated data generation and experimental data collection. (a) Diagram of experiment setup in SSRF. (b) SDD in allpix squared simulator.



(a) (b)

Fig. 6: Evaluation of count-rate estimation performance using the OCR/ICR ratio. (a) OCR/ICR ratio of Cu. (b) OCR/ICR ratio of Co. Deep learning-based methods exhibit a more linear counting response compared with Iterative LASSO [15]. Furthermore, the semi-supervised approach consistently outperforms the fully supervised method, regardless of whether AttnUNet++ [24] or NoA-DNN [25] is used as the backbone.

probability assumptions based on a Poisson process [25], which leads to more stable estimation under low pile-up conditions.

In contrast, at high count-rate conditions, such physical-model-based assumptions become less sufficient to capture the complex pile-up patterns present in real measurements. In this condition, the more expressive architecture of AttnUNet++ enables richer representation learning and therefore yields better performance in modeling high count-rate signals.

These results further suggest that the proposed semi-supervised learning framework provides a solid foundation that enables complex deep learning architectures to fully exploit their modeling capacity in challenging experimental conditions.

D. Spectrum Estimation

In this part, we compare the spectrum estimation performance of three baseline methods: Iterative LASSO, AttnUNet++, and PHE [21] (spectrum estimation–specific method). We also construct two semi-supervised variants by integrating our semi-supervised framework into AttnUNet++ and PHE, denoted as SSL-AttnUNet++ and SSL-PHE.

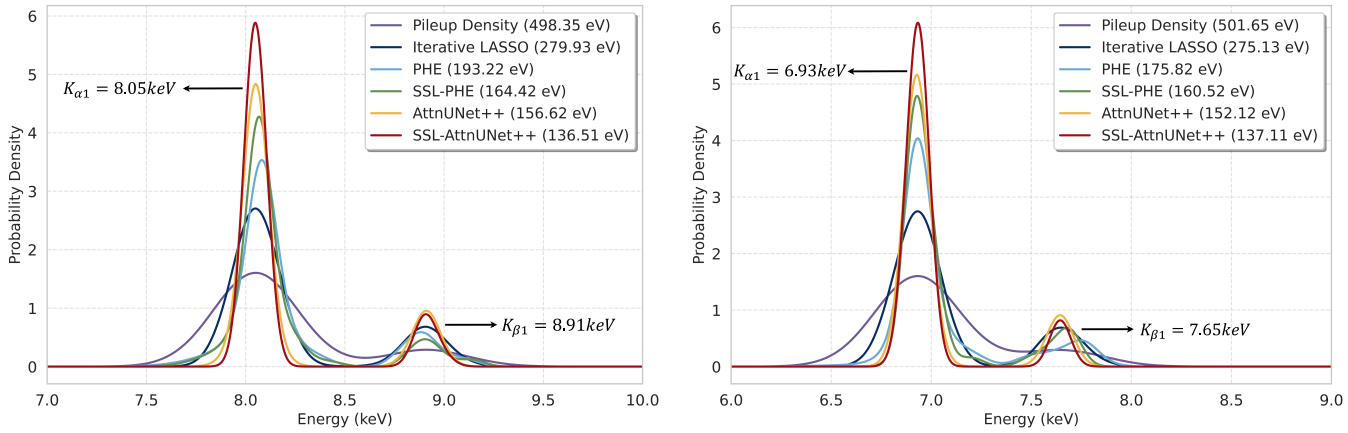
The experiments are conducted under high count-rate conditions, specifically 1142 kcps for Cu and 1115 kcps for Co. As shown in Fig. 7, all methods are able to reconstruct the overall energy distribution from time-domain signals and correctly identify the positions of characteristic emission peaks, while achieving different levels of spectral resolution improvement.

Overall, consistent with the count rate estimation results, deep learning-based methods outperform the classical Iterative LASSO approach. In addition, AttnUNet++ exhibits superior performance compared with PHE, indicating its stronger capability in spectrum reconstruction. Most importantly, across different backbone networks (AttnUNet++ and PHE), the proposed semi-supervised framework consistently improves spectrum reconstruction performance compared with their fully supervised counterparts.

Quantitatively, taking Cu as an example, when using AttnUNet++ as the backbone, the proposed method reduces the FWHM from 156.62 eV to 136.51 eV. When using PHE

TABLE III: Quantitative comparison of counting performance in terms of OCR and ICR (%) across different ICR levels and backbone networks.

Dataset: ICR (kcps)	Iterative LASSO	NoA-DNN	SSL-NoA-DNN	AttnUNet++	SSL-AttnUNet++
Cu: 88	94.31%	98.86%	**99.81%**	95.45%	98.86%
Cu: 458	93.66%	98.03%	**99.12%**	95.41%	98.69%
Cu: 1142	82.92%	93.34%	97.72%	87.00%	**98.43%**
Co: 81	92.59%	99.51%	**98.77%**	95.06%	98.76%
Co: 459	90.41%	98.47%	98.26%	93.03%	**98.47%**
Co: 1115	81.79%	91.84%	95.78%	87.40%	**98.57%**



(a) (b)

Fig. 7: Evaluation of spectral resolution using FWHM. (a) Energy distribution of Cu. (b) Energy distribution of Co. Deep learning-based methods exhibit a higher resolution compared with Iterative LASSO [15]. Furthermore, the semi-supervised approach consistently outperforms the fully supervised method, regardless of whether Attn-UNet++ [24] or PHE [21] is used as the backbone.

as the backbone, the FWHM is reduced from 193.22 eV to 164.42 eV. These results demonstrate that under high count-rate conditions, the proposed method achieves more accurate photon energy estimation and consistently improves spectral resolution.

E. Quantitative Analysis of the Simulation-to-Real Gap

In this part, we characterize the simulation-to-real domain gap and evaluate the effectiveness of the proposed method in bridging this gap. Specifically, we analyze the performance discrepancy between models trained on simulated data and their generalization to real measurements from two perspectives: energy resolution (FWHM) and system throughput efficiency (OCR/ICR).

we design four experimental configurations: (1) *Sim→Sim*, where the model is both trained and evaluated on simulated data; (2) *Sim→Real*, where a model trained solely on simulated data is directly applied to real measurements; (3) *Real→Real* (low count-rate), where both training and evaluation are conducted on low count-rate real data to approximate the empirical upper bound due to the lack of high count-rate labeled data; and (4) the *proposed method*, which jointly leverages simulated data and real unlabeled data within a semi-supervised learning (SSL) framework. Configurations (1), (2), and (4) are trained and evaluated under high count-rate conditions (approximately 1 Mcps), while configuration (3) is conducted at a low count rate of 50 kcps. Quantitative results for all configurations are summarized in Table IV.

Several key observations can be drawn from the quantitative results. First, a clear domain gap is observed when transferring from simulated to real data. Specifically, performance degrades from 132.7 eV and 99.13% (Sim→Sim) to 156.6 eV and 87.04% (Sim→Real), demonstrating that models trained solely on simulated data exhibit limited generalization to real-world measurement conditions.

Second, the Real→Real (low count-rate) configuration achieves 130.8 eV and 99.32% OCR/ICR, providing a practical upper bound for real-domain performance under minimal pile-up conditions.

Third, the proposed method substantially mitigates the simulation-to-real gap. Compared with the Sim→Real baseline, the FWHM is reduced from 156.6 eV to 136.5 eV, while OCR/ICR improves from 87.04% to 98.42%. These results confirm that incorporating real unlabeled data through semi-supervised learning effectively enhances the model's adaptability to real data distributions.

F. Detailed Time-Domain Analysis

To better understand the sources of improvement, a detailed analysis of the underlying time-domain signals was conducted.

TABLE IV: Quantitative evaluation of the sim-to-real gap and the improvement achieved by the proposed method.

Method	Train Data	Test Data	FWHM (eV)↓	OCR/ICR↑
Sim → Sim	Sim	Sim	132.7	99.13%
Sim → Real	Sim	Real	156.6	87.04%
Real → Real	Real (Low count rate)	Real (Low count rate)	130.8	99.32%[1]
Proposed	Sim + Real (SSL)	Real	136.5	98.42%

[1]Cannot be compared directly to other results due to significantly lower count rate.

As shown in Fig. 8, we present representative pulses that are challenging for the fully supervised method (blue), along with the corresponding improved results obtained by the proposed method (green). The red markers indicate the predicted event locations.

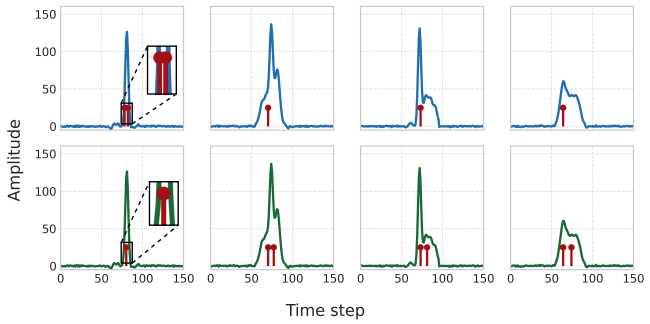


Fig. 8: Case study on representative challenging pulses comparing fully supervised and semi-supervised models. The top row (blue) shows predictions from the fully supervised model, while the bottom row (green) shows results from the semi-supervised model.

The first column corresponds to single-photon events, whereas the remaining columns represent pile-up pulses. Among these challenging cases, single-photon events often exhibit abnormally sharp peaks that deviate from the ideal Gaussian shape. In contrast, pile-up pulses display more complex waveform patterns, posing additional challenges for accurate prediction.

These observations suggest that training exclusively on simulated data limits the model's ability to generalize to the full variability of real pulse shapes. By incorporating high count-rate experimental data into model training, the model is exposed to a broader distribution of pulse characteristics, including subtle variations in amplitude, width, and temporal overlap, thereby improving both count rate accuracy and spectral resolution.

G. Convergence Behavior over Semi-Supervised Iterations

We further investigated how the performance of the proposed framework evolves over multiple iterations of semi-supervised learning. Before discussing the results, it is worth recalling that, as described in Section III-A, the first iteration trains the model exclusively on the labeled dataset, while subsequent iterations progressively incorporate the unlabeled dataset through the semi-supervised pseudo-labeling process.

Fig. 9a illustrates the evolution of the OCR across iterations at ICRs of 88 kcps, 458 kcps, and 1142 kcps. At 88 kcps, the labeled dataset alone sufficiently covers the diversity of pulse shapes in the test dataset, enabling the model to achieve near-optimal OCR after a single iteration. In contrast, at higher count-rates (458 kcps and 1142 kcps), pulse overlap and waveform variability increase substantially, requiring approximately three to five iterations to reach comparable OCR performance. This trend highlights the critical role of iteratively incorporating real experimental data for adapting the model to high count-rate measurement conditions.

Fig. 9b shows the evolution of the energy spectrum resolution for Cu at 1142 kcps. The FWHM decreases steadily with each iteration, indicating continuous improvement in spectral resolution. Notably, the second iteration yields the largest reduction in FWHM, corresponding to the model's first exposure to unlabeled data. Subsequent iterations continue to refine the model, though with diminishing returns, as the most informative pulse patterns are already captured in earlier stages. We observe that the energy spectrum stabilizes after the 4th or 5th iteration. These iterative improvements indicate that the model becomes increasingly capable of representing complex pulse characteristics as more real high count-rate data are incorporated.

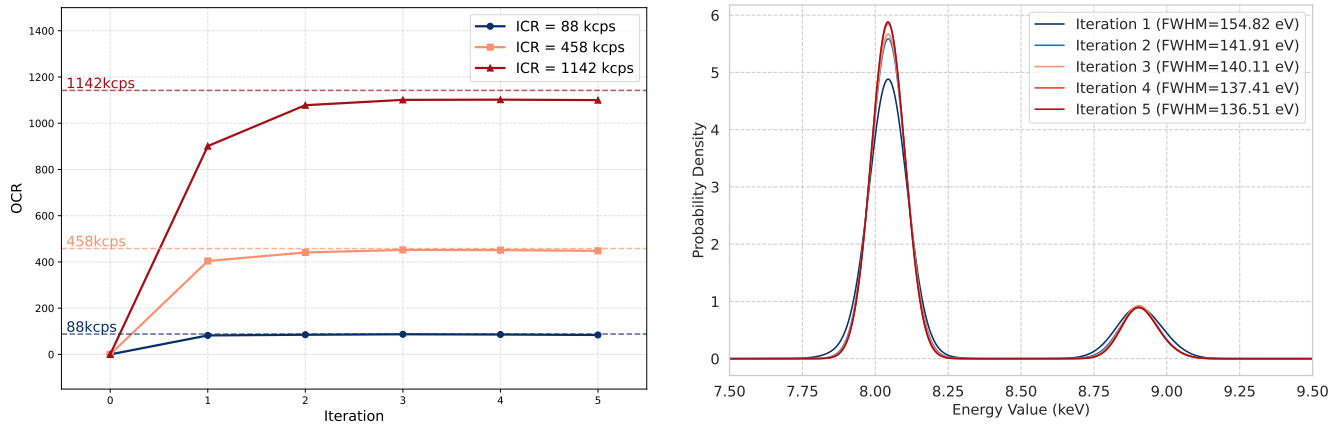
H. Ablation Study

To assess the contribution of each component in our semi-supervised framework, an ablation study was conducted on Cu at an input count rate of 1142 kcps, as summarized in Table V. The Baseline denotes the complete framework, while A1–A4 correspond to variants with specific modules removed or modified.

Removing the pseudo-labeling mechanism (i.e., training only on the labeled dataset, A1) leads to a notable decrease in OCR (-11.6%), accompanied by a substantial degradation in FWHM (+15.09%). This result indicates that unlabeled data play a crucial role in enhancing both count-rate estimation and spectrum estimation. When pseudo-label filtering is omitted (i.e., all pseudo-labels are merged directly into the dataset without reliability screening, A2), performance on count rate accuracy drops even more sharply (-19.9%), demonstrating that filtering unreliable pseudo-labels is essential for preventing error propagation.

Disabling the KL-divergence regularization (A3) slightly reduces OCR (-2.28%) and increases FWHM (+3.96%). Replacing FL with CE during teacher model training further decreases OCR by 1.95% and considerably degrades spectral resolution by 8.86%.

Although KL regularization appears to have only a minor impact on OCR and FWHM, it plays a crucial role in stabilizing the learning process. As illustrated in Fig. 10, removing



(a) (b)

Fig. 9: Convergence Behavior over training iterations for (a) OCR convergence and (b) spectrum convergence.

TABLE V: Ablation study on the proposed semi-supervised learning framework. Metrics include counting performance (OCR, cps) and energy resolution (FWHM, eV). Relative improvements over the baseline are also reported.

ID	Variant	OCR (kcps)↑	OCR vs. Baseline ↑	FWHM (eV)↓	FWHM vs. Baseline ↓
Baseline	Full framework	1123.8	100.00%	136.5	100.00%
A1	w/o pseudo-labels	993.9	88.43%	157.1	115.09%
A2	w/o pseudo-label filtering	900.4	80.09%	124.7	91.36%
A3	w/o KL-divergence	1098.2	97.72%	141.9	103.96%
A4	Teacher: CE instead of FL	1102.5	98.05%	148.6	108.86%

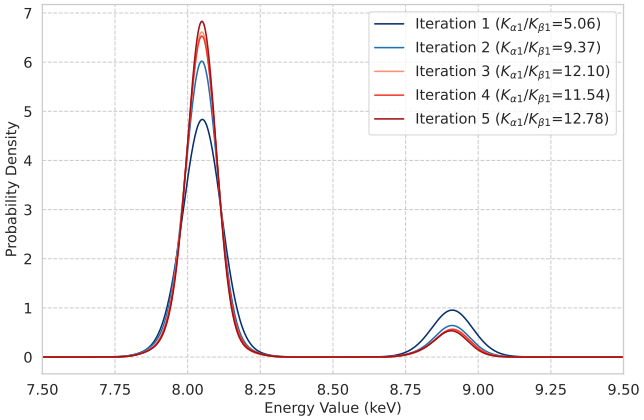


Fig. 10: Spectrum evolution across semi-supervised training iterations without KL regularization. The $K_{\alpha1}/K_{\beta1}$ intensity ratio increases and eventually exceeds the reasonable range.

the KL constraint results in divergent spectral behavior: while the resolution of the $K_{\alpha1}$ peak improves over iterations, the $K_{\beta1}$ peak degrades and may even disappear. Consequently, the $K_{\alpha1}/K_{\beta1}$ intensity ratio exceeds 12.78, deviating substantially from the expected value of approximately 6.18 [40], [41]. This phenomenon indicates that the KL term is essential for maintaining a balanced label distribution and preventing overconfident predictions during semi-supervised training.

To better visualize and interpret the contribution of FL, we further analyze the results using confusion matrices and macro-F_1 scores on simulated data. Specifically, a signal simulator [42], [43] was used to conduct experiments on ^{60}Co. Since this section focuses on identifying general patterns independent of specific implementation configurations, we introduce the dimensionless normalized intensity ρ to characterize the radiation intensity level,

$$\rho = \lambda w, \tag{20}$$

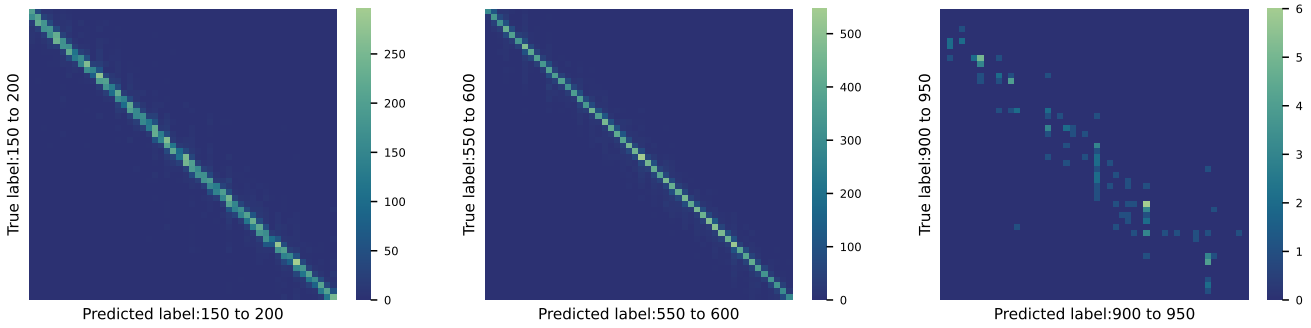
where λ denotes the count rate and w represents the pulse width. This formulation enables direct conversion to system-specific parameters in practical radiation measurement setups.

Fig. 11 and Fig. 12 present the confusion matrices for the low-, mid-, and high-energy regions at $\rho = 0.09$. It can be observed that both FL and CE perform well in the mid-energy region, while their performance degrades in the low- and high-energy regions.

This behavior can be attributed to the intrinsic characteristics of the nuclear energy spectrum, as illustrated in Figure 12.20 (p. 442) of [35]. First, the energy distribution of ^{60}Co exhibits a significant class imbalance. The mid-energy region has a more concentrated spectral distribution, making it easier to classify. In contrast, the high-energy region contains fewer samples due to limited statistics. Second, the low-energy region suffers from low SNR and is heavily affected by noise, which further increases the difficulty of accurate prediction.

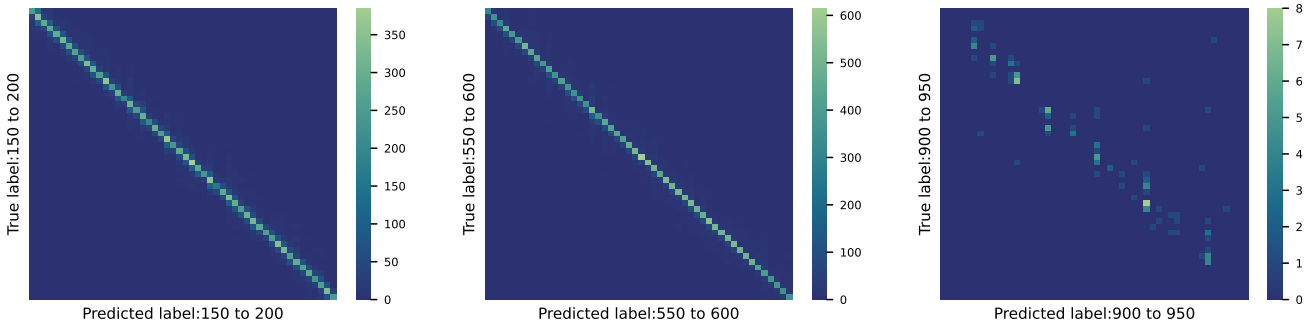
Nevertheless, FL consistently outperforms CE in these challenging regions, while their performance remains comparable in the mid-energy range. These results demonstrate that FL provides improved robustness for minority and hard-to-distinguish samples.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2026.3706156



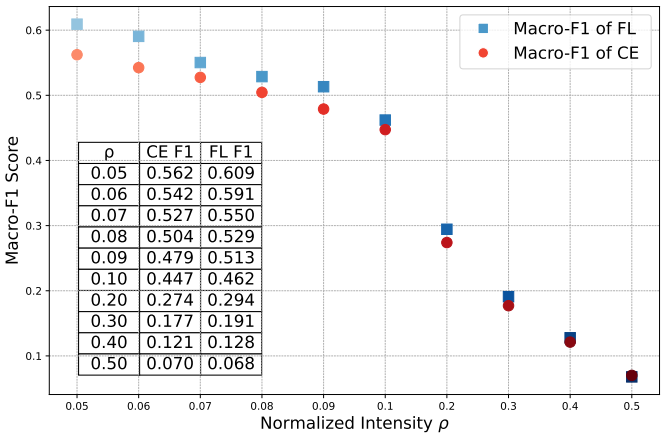
(a) (b) (c)

Fig. 11: Confusion matrix across different energy regions when the model is trained with CE. (a) Class 150–200 (low-energy region). (b) Class 550–600 (mid-energy region). (c) Class 900–950 (high-energy region).



(a) (b) (c)

Fig. 12: Confusion matrix across different energy regions when the model is trained with FL. (a) Class 150–200 (low-energy region). (b) Class 550–600 (mid-energy region). (c) Class 900–950 (high-energy region).



ρ	CE F1	FL F1
0.05	0.562	0.609
0.06	0.542	0.591
0.07	0.527	0.550
0.08	0.504	0.529
0.09	0.479	0.513
0.10	0.447	0.462
0.20	0.274	0.294
0.30	0.177	0.191
0.40	0.121	0.128
0.50	0.070	0.068

Fig. 13: Macro-F_1 comparison between CE and FL across different intensity levels.

Fig. 13 presents the Macro-F_1 evaluation across different normalized intensities ρ. When ρ is below 0.5, FL consistently outperforms CE. As the intensity increases, the performance gap gradually narrows, and when ρ exceeds 0.5, CE slightly surpasses FL. This behavior can be attributed to the severe pile-up effect at extrmmely high count rates, which makes all classes difficult to discriminate. In this condition, FL's down-weighting of easy samples may inadvertently reduce overall performance.

Based on these findings, we recommend applying FL when ρ is below 0.5, while standard CE is preferable when ρ exceeds 0.5. For the radiation measurement setups discussed in Section IV-A, given $w = 180\,\text{ns}$, the threshold $\rho = 0.5$ corresponds to a count rate of $\lambda = 2.78\,\text{Mcps}$.

I. Computational Efficiency and Practical Applicability

In this part, we evaluate the computational efficiency and practical applicability of the proposed framework across different backbones, namely AttnUNet++, NoA-DNN, and PHE.

We first analyze model complexity and training cost. All models are trained under identical experimental conditions as specified in Section IV-A. Computational complexity is reported in terms of floating-point operations (FLOPs), together with the corresponding training time.

To assess practical applicability, inference latency is measured on two representative hardware platforms: a server-grade GPU (NVIDIA A30) and an embedded GPU (NVIDIA Jetson Orin Nano). The A30 environment is configured with Ubuntu 22.04 and PyTorch 2.8 (CUDA 12.6) using Python 3.10, while the Jetson platform uses JetPack 6.2.1 (CUDA 12.6) and PyTorch 2.5, also with Python 3.10. Inference is performed with a batch size of 128 on both platforms.

TABLE VI: Computational complexity, training time, and inference latency of different backbones on two hardware platforms.

Method	FLOPs (M)	Training Time (h)	A30 Latency (μs)	Jetson Latency (μs)
SSL-AttnUnet++	15,128.33	19.81	991.02	12,759.21
SSL-NoA-DNN	0.16	0.20	1.41	7.03
SSL-PHE	0.60	1.44	1.95	15.23

In addition, `torch.compile` [44] is used for graph-level optimization to reduce inference overhead. All reported latency values correspond to the per-sample inference time for an input signal comprising 1024 sampling points. The quantitative results are summarized in Table VI.

As shown in the table, substantial variations in both training time and inference latency are observed across backbone networks. These differences are directly attributable to their respective computational complexities, as reflected by FLOPs. Specifically, SSL-AttnUNet++ incurs the highest computational cost, resulting in a significantly longer training time and higher inference latency. However, as discussed in Section IV-C and Section IV-D, SSL-AttnUNet++ generally achieves superior performance in both count rate estimation and spectrum reconstruction. This is the well-known trade-off between accuracy and computational cost in deep learning models.

To further assess whether the models can achieve real-time processing, inference latency is interpreted in the context of the data acquisition pipeline. Given a sampling rate of 250 MHz and a digital pulse processing (DPP) scheme with an effective downsampling factor of 2, each sampling point corresponds to a temporal resolution of 8 ns. Consequently, an input signal of 1024 sampling points spans approximately 8.192 μs, which defines the real-time inference deadline for single-pulse processing.

Under this criterion, SSL-NoA-DNN satisfies the real-time requirement on both platforms (1.41 μs, 7.03 μs). SSL-PHE meets the constraint on the server-grade platform, but fails to meet it on the embedded device (1.95 μs, 15.23 μs). SSL-AttnUNet++, however, does not satisfy the real-time requirement on either platform (991.02 μs, 12759.21 μs), making it more suitable for offline analysis scenarios where accuracy is prioritized over latency.

Finally, it is worth emphasizing that the proposed framework is fully compatible with standard model compression techniques, such as quantization and structured pruning. This provides a clear pathway for further reducing computational overhead and enabling deployment on resource-constrained edge devices.

J. Parameter Selection and Analysis

In this section, we discuss the selection of parameters listed in Table II, covering the semi-supervised training strategy, loss function design, and dataset composition.

We first consider the number of semi-supervised iterations, which is set to 5. As shown in Fig. 9, performance improvements saturate after 4–5 iterations, indicating that the model has nearly converged. In practice, early stopping based on validation metrics is recommended, as excessive iterations may lead to overfitting to noisy pseudo-labels and consequently degrade generalization performance.

We next analyze the loss function design. The Focal Loss focusing parameter $\gamma = 1$ and the KL-divergence weighting coefficient $\alpha = 0.1$ are selected via grid search to jointly optimize OCR and FWHM. Larger values of γ tend to overemphasize noisy outliers, whereas smaller values reduce the loss to standard CE, thereby weakening its ability to address class imbalance. Ablation studies (Table V) further demonstrate that removing the KL-divergence term degrades both reconstruction quality and training stability. However, excessively large values of α overly constrain the student model, limiting its ability to effectively exploit unlabeled data.

The dataset consists of 5,000 simulated samples, 5,000 experimental samples, and 20,000 unlabeled samples. The large pool of unlabeled data plays a critical role in improving distribution coverage and model robustness. In general, increasing the amount of unlabeled data improves generalization performance, provided that sufficient computational resources are available.

Based on the data characteristics and physical constraints, the pseudo-label threshold $N_{\max}$ is set to 3. Given the experimentally measured average pulse rise time $t_r = 90$ ns and the maximum count rate of approximately $\lambda = 10^6$ cps, $N_{\max}$ is computed according to Eq. (11) and Eq. (12).

Finally, the remaining hyperparameters not included in Table II (e.g., network depth, hidden dimension, learning rate, and number of training epochs) are inherited from the backbone network. Since the proposed framework is architecture-agnostic, these settings follow standard configurations of the chosen backbone and are not further tuned in this study, ensuring a fair evaluation of the proposed method.

V. CONCLUSIONS

This work proposes a semi-supervised learning framework to address pulse pile-up in nuclear spectroscopy under high count-rate conditions, where the lack of reliable labels limits its fully supervised approaches. By integrating pseudo-label generation, physics-guided refinement, and a stage-wise loss design, the proposed framework achieves improved performance in both spectral resolution and count-rate estimation. Ablation studies further confirm the critical roles of pseudo-label generation and refinement, KL regularization, and focal loss.

In future work, we are particularly interested in exploring self-supervised and few-shot learning methods for energy spectrum reconstruction in scenarios with limited labeled data. In addition, since different nuclear radiation detectors exhibit distinct pulse characteristics, we aim to investigate transfer learning approaches to extend the proposed method to a broader range of real-world detection applications.

REFERENCES

[1] R. Jenkins *et al.*, *X-ray fluorescence spectrometry*, vol. 2. Wiley Online Library, 1999.

[2] R. Van Grieken and A. Markowicz, *Handbook of X-ray Spectrometry*. CRC press, 2001.

[3] M. Nakhostin, Z. Podolyak, P. Regan, and P. Walker, "A digital method for separation and reconstruction of pile-up events in germanium detectors," *Review of Scientific Instruments*, vol. 81, no. 10, 2010.

[4] C. Imperiale and A. Imperiale, "On nuclear spectrometry pulses digital shaping and processing," *Measurement*, vol. 30, no. 1, pp. 49–73, 2001.

[5] M.-R. Mohammadian-Behbahani and S. Saramad, "A comparison study of the pile-up correction algorithms," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 951, p. 163013, 2020.

[6] C. Williams, "Reducing pulse height spectral distortion by means of dc restoration and pile-up rejection," *IEEE Transactions on Nuclear Science*, vol. 15, no. 1, pp. 297–302, 2007.

[7] S. Rozen, "Pile up rejection cicuits," *Nuclear Instruments and Methods*, vol. 11, pp. 316–320, 1961.

[8] M. Bolic, V. Drndarevic, and W. Gueaieb, "Pileup correction algorithms for very-high-count-rate gamma-ray spectrometry with nai (tl) detectors," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 1, pp. 122–130, 2009.

[9] P. A. Scoullar and R. J. Evans, "Maximum likelihood estimation techniques for high rate, high throughput digital pulse processing," in *2008 IEEE Nuclear Science Symposium Conference Record*, pp. 1668–1672, IEEE, 2008.

[10] J. Liu, H. Li, Y. Wang, S. Kim, Y. Zhang, S. Liu, H. Baghaei, R. Ramirez, and W.-H. Wong, "Real time digital implementation of the high-yield-pileup-event-recover (hyper) method," in *2007 IEEE Nuclear Science Symposium Conference Record*, vol. 6, pp. 4230–4232, IEEE, 2007.

[11] W.-H. Wong and H. Li, "A scintillation detector signal processing technique with active pileup prevention for extending scintillation count rates," *IEEE Transactions on Nuclear Science*, vol. 45, no. 3, pp. 838–842, 1998.

[12] X. Wen and H. Yang, "Study on a digital pulse processing algorithm based on template-matching for high-throughput spectroscopy," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 784, pp. 269–273, 2015.

[13] M.-R. Mohammadian-Behbahani and S. Saramad, "Pile-up correction algorithm based on successive integration for high count rate medical imaging and radiation spectroscopy," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 897, pp. 1–7, 2018.

[14] X. Wang, Q. Xie, Y. Chen, M. Niu, and P. Xiao, "Advantages of digitally sampling scintillation pulses in pileup processing in pet," *IEEE Transactions on Nuclear Science*, vol. 59, no. 3, pp. 498–506, 2012.

[15] T. Trigano, I. Gildin, and Y. Sepulcre, "Pileup correction algorithm using an iterated sparse reconstruction method," *IEEE Signal Processing Letters*, vol. 22, no. 9, pp. 1392–1395, 2015.

[16] B. Liu, M. Liu, M. He, Y. Ma, and X. Tuo, "Model-based pileup events correction via kalman-filter tunnels," *IEEE Transactions on Nuclear Science*, vol. 66, no. 1, pp. 528–535, 2018.

[17] T. Trigano, A. Souloumiac, T. Montagu, F. Roueff, and E. Moulines, "Statistical pileup correction method for hpge detectors," *IEEE Transactions on Signal Processing*, vol. 55, no. 10, pp. 4871–4881, 2007.

[18] T. Trigano, E. Barat, T. Dautremer, and T. Montagu, "Fast digital filtering of spectrometric data for pile-up correction," *IEEE Signal Processing Letters*, vol. 22, no. 7, pp. 973–977, 2014.

[19] C. Mclean, M. Pauley, and J. H. Manton, "Non-parametric decompounding of pulse pile-up under gaussian noise with finite data sets," *IEEE Transactions on Signal Processing*, vol. 68, pp. 2114–2127, 2020.

[20] P. Ilhe, É. Moulines, F. Roueff, and A. Souloumiac, "Nonparametric estimation of mark's distribution of an exponential shot-noise process," *Electronic Journal of Statistics*, vol. 9, no. 2, pp. 3098 – 3123, 2015.

[21] B. Jeon, S. Lim, E. Lee, Y.-S. Hwang, K.-J. Chung, and M. Moon, "Deep learning-based pulse height estimation for separation of pile-up pulses from nai (tl) detector," *IEEE Transactions on Nuclear Science*, vol. 69, no. 6, pp. 1344–1351, 2021.

[22] M. Kamuda, J. Zhao, and K. Huff, "A comparison of machine learning methods for automated gamma-ray spectroscopy," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 954, p. 161385, 2020.

[23] T. Trigano and D. Bykhovsky, "Deep learning based method for activity estimation from short-duration gamma spectroscopy recordings," *IEEE Transactions on Instrumentation and Measurement*, 2024.

[24] Y. Huang, C. Lin, D. Bykhovsky, T. Trigano, Z. Chen, X. Zheng, and Y. Zhu, "Deep learning based energy spectrum estimation for high counting rate nuclear spectrometry," *IEEE Transactions on Instrumentation and Measurement*, 2025.

[25] I. Morad, M. Ghelman, D. Ginzburg, A. Osovizky, and N. Shlezinger, "Model-based deep learning algorithm for detection and classification at high event rates," *IEEE Transactions on Nuclear Science*, vol. 71, no. 5, pp. 970–980, 2024.

[26] B. Sadeghi, D. Scriven, A. Chester, R. Fox, S. Liddick, and G. Cerizza, "A machine learning framework for accurate and robust analysis of radiation detector pulses," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, p. 170971, 2025.

[27] V. Daniel Elvira, "Impact of detector simulation in particle physics collider experiments," *Physics Reports*, vol. 695, pp. 1–54, 2017. Impact of Detector Simulation in Particle Physics Collider Experiments.

[28] A. Wilkinson, R. Radev, and S. Alonso-Monsalve, "Contrastive learning for robust representations of neutrino data," *Physical Review D*, vol. 111, no. 9, p. 092011, 2025.

[29] Y. Huang, X. Zheng, Y. Zhu, T. Trigano, D. Bykhovsky, and Z. Chen, "Deep learning based pile-up correction algorithm for spectrometric data under high-count-rate measurements," *Sensors*, vol. 25, no. 5, p. 1464, 2025.

[30] R. Xie, C. Wang, W. Zeng, and Y. Wang, "An empirical study of the collapsing problem in semi-supervised 2d human pose estimation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 11240–11249, 2021.

[31] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988, 2017.

[32] Goldberger, Gordon, and Greenspan, "An efficient image similarity measure based on approximations of kl-divergence between two gaussian mixtures," in *Proceedings Ninth IEEE International conference on computer vision*, pp. 487–493, IEEE, 2003.

[33] F. Y. Daniel and J. A. Fessler, "Mean and variance of coincidence counting with deadtime," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 488, no. 1-2, pp. 362–374, 2002.

[34] D.-C. Li, Z.-G. Ren, L. Yang, Z. Qi, X.-T. Meng, and B.-T. Hu, "A novel acquisition method for nuclear spectra based on pulse area analysis," *Chinese Physics C*, vol. 39, no. 4, p. 046201, 2015.

[35] G. F. Knoll, *Radiation detection and measurement*. John Wiley & Sons, 2010.

[36] S. Agostinelli, J. Allison, K. a. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. Barrand, *et al.*, "Geant4—a simulation toolkit," *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250–303, 2003.

[37] S. Spannagel, K. Wolters, D. Hynds, N. A. Tehrani, M. Benoit, D. Dannheim, N. Gauvin, A. Nürnberg, P. Schütze, and M. Vicente, "Allpix2: A modular simulation framework for silicon detectors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 901, pp. 164–172, 2018.

[38] G. Aad, X. S. Anduaga, S. Antonelli, M. Bendel, B. Breiler, F. Castrovillari, J. Civera, T. Del Prete, M. T. Dova, S. Duffin, *et al.*, "The atlas experiment at the cern large hadron collider," 2008.

[39] B. Bergmann, T. Billoud, C. Leroy, and S. Pospisil, "Characterization of the radiation field in the atlas experiment with timepix detectors," *IEEE Transactions on Nuclear Science*, vol. 66, no. 7, pp. 1861–1869, 2019.

[40] R. Yılmaz, "Kβ/kα x-ray intensity ratios for some elements in the atomic number range $28 \leq z \leq 39$ at 16.896 kev," *Journal of Radiation Research and Applied Sciences*, vol. 10, no. 3, pp. 172–177, 2017.

[41] G. Hölzer, M. Fritsch, M. Deutsch, J. Härtwig, and E. Förster, "K α 1, 2 and k β 1, 3 x-ray emission lines of the 3 d transition metals," *Physical Review A*, vol. 56, no. 6, p. 4554, 1997.

[42] Z. Chen, D. Bykhovsky, X. Zheng, T. Trigano, and Y. Zhu, "Gasim: A python class to generate simulated time signals for gamma spectroscopy," *SoftwareX*, vol. 29, p. 102037, 2025.

[43] D. Bykhovsky, Z. Chen, Y. Huang, X. Zheng, and T. Trigano, "Advanced spectroscopy time-domain signal simulator for the development of machine and deep learning algorithms," *IEEE Sensors Letters*, 2025.

[44] PyTorch Team, "Introduction to torch.compile." https://docs.pytorch.org/tutorials/intermediate/torch_compile_tutorial.html, 2023. Accessed: 2026-04-16.