

FhyMetric-Net: Interpretable Mixed Radioisotope Identification Model Integrating Prior Characteristic Peak Physical Information and Feature Metric Constraints

Authors: Zhang, Dr. CaoLin, Prof. Jiangmei Zhang, Liu, Dr. Haolin, Yang, Dr. Guowei, Wang, Dr. JiaQi, rui, Mr. tang, Zhang, Dr. CaoLin

Date: 2025-06-25T00:00:00+00:00

Abstract

Automatic identification of radioactive isotopes through energy spectrum analysis is vital for remote, unmanned monitoring of radioactive contamination and rapid early warning. In recent years, deep learning methods have advanced significantly, outperforming traditional approaches in recognition accuracy. However, their purely data-driven nature and the “black-box” characteristics of neural networks result in poor interpretability, a high risk of overfitting, and uncontrollable errors, limiting their use in high-reliability fields like the nuclear industry. We present FhyMetric-Net, a novel multi-label classification model that integrates physical constraints with data-driven techniques. This model automatically infers the probability of mixed nuclides and provides weight interpretations consistent with expert knowledge. Our approach is groundbreaking in embedding prior characteristic peak physical information into neural networks, effectively constraining the feature weight optimization space for improved reliability and interpretability. We also introduce a novel metric constraint method in the feature space, tailored for mixed nuclide samples, which enhances the model’s ability to extract discriminative features. By establishing a clear causal link between predicted probabilities and channel addresses, FhyMetric-Net overcomes the interpretability challenges of traditional dense fully connected layers. We conducted more challenging quantitative tests than previous studies. When faced with challenges such as an increased number of mixed radionuclides, variations in Gaussian broadening coefficients, and differences in detector types, the proposed model consistently maintained an F1 score above 95%, achieving state-of-the-art (SOTA) performance, while the model’s parameter count was only 1.58% of the ResNet-18 model. In scenarios with low gross count and low signal-to-noise ratio (SNR), its overall performance also demonstrated signifi-

cant advantages. Qualitative analysis further confirmed the model' s strong physical interpretability. This achievement will advance the application of automated mixed radionuclide identification technology in high-reliability fields of the nuclear industry.

Full Text

Preamble

FhyMetric-Net: Interpretable mixed radioisotope identification model integrating prior characteristic peak physical information and feature metric constraints*

Cao-Lin Zhang,^{1, 2} Jiang-Mei Zhang,^{1, 2}, † Hao-Lin Liu,² Guo-Wei Yang,² Jia-Qi Wang,^{1, 2} and Rui Tang¹ ¹School of Information Engineering, Southwest University of Science and Technology, Mianyang 621010, China ²CAEA Innovation Center of Nuclear Environmental Safety Technology, Southwest University of Science and Technology, Mianyang 621010, China

Automatic identification of radioactive isotopes through energy spectrum analysis is vital for remote, unmanned monitoring of radioactive contamination and rapid early warning. In recent years, deep learning methods have advanced significantly, outperforming traditional approaches in recognition accuracy. However, their purely data-driven nature and the “black-box” characteristics of neural networks result in poor interpretability, a high risk of overfitting, and uncontrollable errors, limiting their use in high-reliability fields like the nuclear industry. We present FhyMetric-Net, a novel multi-label classification model that integrates physical constraints with data-driven techniques. This model automatically infers the probability of mixed nuclides and provides weight interpretations consistent with expert knowledge. Our approach is groundbreaking in embedding prior characteristic peak physical information into neural networks, effectively constraining the feature weight optimization space for improved reliability and interpretability. We also introduce a novel metric constraint method in the feature space, tailored for mixed nuclide samples, which enhances the model' s ability to extract discriminative features. By establishing a clear causal link between predicted probabilities and channel addresses, FhyMetric-Net overcomes the interpretability challenges of traditional dense fully connected layers.

We conducted more challenging quantitative tests than previous studies. When faced with challenges such as an increased number of mixed radionuclides, variations in Gaussian broadening coefficients, and differences in detector types, the proposed model consistently maintained an F1 score above 95%, achieving state-of-the-art (SOTA) performance, while the model' s parameter count was only 1.58% of the ResNet-18 model. In scenarios with low gross count and low signal-to-noise ratio (SNR), its overall performance also demonstrated significant advantages. Qualitative analysis further confirmed the model' s strong physical interpretability. This achievement will advance the application of au-

tomated mixed radionuclide identification technology in high-reliability fields of the nuclear industry.

Keywords: Radioisotope Identification, Deep Learning, Physical Interpretability, Feature Metric Constraints, Interdisciplinary

Introduction

Radioisotope identification (RIID) is a key technology for detecting radioactive contamination and identifying radioactive sources, which plays a crucial role in nuclear safety and environmental protection [1–5].

Traditional radioisotope identification (RIID) methods mainly search for characteristic peaks based on derivative, Fourier transform or wavelet analysis techniques, and match them with a predefined nuclide library (containing key information such as energy and branching ratio) to determine the type of nuclide [6–10]. However, due to factors such as statistical fluctuations and peak overlapping, these methods typically require manual iterative adjustment of parameters to achieve optimal noise smoothing and peak finding performance. This process is not only cumbersome, but also susceptible to human factors, which limits the efficiency and accuracy of identification [11].

In recent years, methods based on Deep Learning (DL), which can automatically adjust parameters to extract deep features from energy spectra, have been widely studied. Existing research results generally indicate that DL models have significant advantages in high identification accuracy when facing challenges such as low count rates and low signal-to-noise (SNR) ratios [12–16]. However, the mainstream DL methods are data-driven approaches based on large-scale training data, which establish an end-to-end mapping between input energy spectra and output probabilities to achieve high identification accuracy. Due to random factors such as temperature, background, and shielding affecting the shape of energy spectra, obtaining large-scale measured samples is challenging. RIID tasks often exhibit the characteristics of “Few Shot + Multi Label”. The “black-box” nature of neural networks (NN) poses risks such as poor interpretability, weak generalization, and uncontrollable errors, limiting their application in the nuclear industry, which demands high reliability [17].

Researchers have studied the interpretability of the NN model to understand the reasons behind the high identification accuracy of NN. Mario Gomez-Fernandez et al. showed the connection between the region of interest (ROI) and the physical characteristics (e.g., photoelectric peaks) using thermal vector maps. Yu Wang et al. [19] used Class Activation Mapping (CAM) to visualize and explain the key regions learned by the model. These findings suggest that an effective classification model should focus on characteristic peak regions in the energy spectrum, rather than unrelated areas like background noise. Thus, establishing a constraint between the model output probabilities and the characteristic peaks in the energy spectrum is crucial for enhancing both classification performance and physical interpretability.

Some researchers have worked to improve the robustness of the model. Zakariya Chaouai et al. [20] improved the robustness of the model by using adversarial learning techniques to reduce the likelihood of NN misguidedness. Hao-Lin Liu et al. [21] performed classification by integrating local and global characteristics of a deep convolutional neural network (CNN), enhancing intraclass similarity and differences between classes. For RIID in urban environments [22], a detection model based on a weighted k-nearest neighbors (KNN) framework was developed to extract discriminative features from energy spectra and minimize inter-class similarity for the same radioactive material. However, these studies remain purely data-driven and fail to fully incorporate physical constraints between characteristic peaks regions and radionuclide presence probabilities, lacking physical interpretability.

Attention has become a key concept in the rapidly advancing field of DL. In computer vision, salient visual attention models have been shown to extract low-level visual features to identify potential key regions [23, 24]. In multi-label image classification tasks, it has been proven to effectively capture semantic and spatial relationships between labels in images [25, 26]. These studies align with RIID, as each category is linked to specific characteristics of the full-energy peak region, such as spatial location and intensity. Jiaqian Sun et al. [27] presented a NN model that combines convolution and self-attention mechanisms. Wang et al. [28] proposed a DL-based method for recognizing multiple radioactive nuclides using a channel attention module. This model explains how it utilizes feature information from the photopeak and Compton edge by interpreting spectral features. However, existing data-driven methods fail to constrain the relationship between output results and characteristic peaks, causing NN to learn irrelevant features like background noise.

To address these issues, we have developed a novel model that combines physical constraints with data-driven approaches for mixed radionuclide identification. The model integrates prior physical information about the radionuclide characteristic peaks with feature space metric constraints, and can effectively capture the latent relationship between radionuclide presence probabilities and characteristic peaks through a learnable convolutional network, offering intuitive physical interpretability.

The specific contributions are outlined as follows: (1) A residual network incorporating multi-scale dilated convolutions was utilized for energy spectrum feature extraction, allowing the model to efficiently integrate both local details and global spatial semantic features. (2) The optimization space of NN weights was constrained by prior information on radionuclide physical characteristic peaks, thereby enhancing the reliability and interpretability of the model. (3) A feature metric constraint method tailored for the multi-label classification (MLC) of mixed radionuclides was proposed, constructing a discriminative feature space. (4) A clear and intuitive computational process was established that links radionuclide prediction probabilities with feature weights, enhancing the causal logic and physical interpretability of the model inference.

[Figure 1: see original paper]

Materials and Methods

A. Backbone Network

To accelerate model convergence and reduce overfitting, a Batch Normalization layer was applied after each convolutional layer, as expressed by the following formula:

$$\text{BN}(x) = \frac{x - E[x]}{\sqrt{\text{Var}[x] + \epsilon}} \cdot \gamma + \beta$$

The overall architecture of the proposed model, shown in Fig. 1, consists of four components: the backbone network, prior physical constraints, feature metric constraints, and MLC.

Deep features of the energy spectrum were extracted using a one-dimensional CNN. Delicate local features need to be extracted for stacked, narrow or weak peaks. However, some radionuclide characteristic peaks, such as those of ^{152}Eu , are distributed across multiple energy ranges, necessitating the extraction of global spatial semantic features.

A novel neural network architecture called Multi-Scale Dilated Residual Network (MSDR-Net) was proposed. This architecture integrates local detail features and global spatial semantic features in the energy spectrum, effectively addressing the vanishing gradient problem while maintaining a lightweight structure. The input is a one-dimensional vector of the gamma energy spectrum, $X \in \mathbb{R}^{1 \times 1024}$. Each radionuclide category was labeled with a binary sequence $P = (P_1, P_2, \dots, P_C)^T \in \mathbb{R}^C$, where $P_i = 1$ if category i is present, and 0 otherwise. Three parallel residual modules then process the input features at multiple scales using different dilation rates. Dilated convolutions expand the receptive field by inserting gaps between kernel elements, without increasing the number of parameters.

The output of the dilated convolution can be expressed as:

$$Y[i] = \sum_{k=1}^K (W[k] \cdot X[i + r \cdot (k - \lfloor K/2 \rfloor)]) + b$$

The receptive field is defined by the following formula:

$$\text{Receptive Field} = (K - 1) \times r + 1$$

Where $Y[i]$ is the output feature, $X[i]$ is the input feature, $W[k]$ is the convolution kernel, b is the bias, $K = 3$ is the kernel size, and $r = 4, 6, 8$ are the

dilation rates. Each module captured contextual information from distinct energy ranges using two layers of dilated convolution and adjusted the channel dimension with a 1×1 convolution.

Leaky ReLU activation function was applied in the paper, which is defined as follows:

$$\sigma(x) = \begin{cases} x, & x \geq 0 \\ kx, & x < 0 \end{cases}$$

where k is a small constant (0.15 in this study) that allows a small gradient to flow through negative inputs, preventing neurons from becoming inactive. It preserves the nonlinear activation characteristics and enhances the stability of gradient propagation.

Each branch output was padded to a fixed length of 1024 using dilated convolutions. The outputs were then concatenated along the channel dimension into 384-dimensional features, which were assigned to the number of classes by a 1×1 convolutional layer, resulting in a class feature $F_{MSDR} \in \mathbb{R}^{C \times 1024}$.

$$F_{MSDR} = f_{MSDR}(X; \theta_{MSDR}), \quad F_{MSDR} \in \mathbb{R}^{C \times 1024}$$

$f_{MSDR}(\cdot)$ represents the parametric function of the backbone network, while θ_{MSDR} refers to the learned parameters, including the convolutional kernel weights w and biases b .

In general, the input energy spectrum samples undergo feature decomposition for each class using MSDR-Net, facilitating subsequent physical information processing and feature space constraints.

B. Prior Physical Constraints

The characteristic peaks energy and full width at half maximum (FWHM) of gamma radionuclides represent physical prior knowledge that the model can incorporate.

Given the output features of the backbone network $F_{MSDR} \in \mathbb{R}^{C \times 1024}$, our objective is to enhance the corresponding full-energy peak features for specific labels locally, while suppressing irrelevant noise interference, even when these features are weak on a global scale.

In this study, we introduced a constant gamma characteristic peaks prior information matrix $M_{PhyPeak}$ to apply attention weighting to F_{MSDR} , generating attention features F_{PhyAtt} for each label.

$$\tilde{F}_{PhyAtt} = M_{PhyPeak} \times F_{MSDR}, \quad \tilde{F} \in \mathbb{R}^{C \times 1024}$$

For each characteristic peak of radioactive isotopes, the region of characteristic peak was assigned a value of 1 based on the FWHM. Other regions were given a small weight of 0.1, preventing the majority of physical information from being disregarded, thus mitigating potential risks of training instability.

Based on the energy calibration, the matrix $M_{PhyPeak}$ indexed by channels was obtained:

$$M_{PhyPeak} = [V_{rad1}, V_{rad2}, \dots, V_{radC}]^T$$

In the above equation, C represents the number of labels. $V_{radn} = [W_1, W_2, \dots, W_N]$ is the weight mask of the region of characteristic peaks for each radionuclide, n is the category label index and $N = 1024$ is the number of input sample channels.

During training, $M_{PhyPeak}$ is a constant attention weight matrix derived from physical prior information. Traditional data-driven methods rely on weakly supervised learning to generate attention weights adaptively. Therefore, the method proposed in this paper imposes stronger constraints.

Adaptive Average Pooling was applied to reduce feature dimensionality and smooth weights. Additionally, a 1×1 convolution operation was employed for further weight fine-tuning.

$$F_{PhyAtt} = \text{Conv}_{1 \times 1}(\text{AdaAvg}_{128}(\tilde{F}_{PhyAtt})), \quad F_{PhyAtt} \in \mathbb{R}^{C \times 128}$$

C. Feature Metric Constraints

Implementing feature space metric constraints to separate dissimilar classes and group similar ones aids the model in extracting discriminative features, thereby improving model reliability, as shown in previous studies [22, 29–31]. However, in MLC tasks involving mixed radionuclides, sample labels often overlap, making it challenging to apply metric constraints at the sample level. In this study, we proposed a novel feature space metric constraint method specifically designed for radionuclide identification tasks.

During training, F_{PhyAtt} represents the feature map of the C categories. $F_{Batch} = [F_{PhyAtt}^{t1}, F_{PhyAtt}^{t2}, \dots, F_{PhyAtt}^{tB}]^T \in \mathbb{R}^{B \times C \times 128}$ is the batch feature output. Based on the corresponding binary label mask $y \in \{0, 1\}^C$, we can obtain the existence feature set $E = \{f_i | y_i = 1\}$ and the absent feature set $A = \{f_j | y_j = 0\}$.

Existence Constraint: For the existence feature set $E = \{f_i | y_i = 1\}$, the distance metric is constructed as D_{exist} :

$$D_{exist} = [d(f_i, f_j)]_{f_i, f_j \in E}, \quad d(a, b) = \|a - b\|_2$$

Define the intra-class mask $M^+ = [y_i = y_j]_{i,j}$ and the inter-class mask $M^- = \neg M^+$. The loss function is based on triplet loss, which aims to minimize the distance between features of the same class while maximizing the distance between features of different classes.

$$L_{triplet} = \frac{1}{|\mathcal{E}|} \sum_{i \in \mathcal{E}} \max(0, D_{exist}(i, p^*) - D_{exist}(i, n^*) + \gamma)$$

$$p^* = \arg \max_{f_j \in M^+} D_{exist}(i, j), \quad n^* = \arg \min_{f_k \in M^-} D_{exist}(i, k)$$

In the above equation, $\mathcal{E}$ is the set of all existing features, p^* is the most similar feature in the same class, and n^* is the most dissimilar feature in a different class. The γ is a margin hyperparameter that controls the minimum distance between intra-class and inter-class features.

Absent Constraint: For the absent feature set $A = \{f_j | y_j = 0\}$, calculate the distance between it and the existing features, and constrain the minimum distance.

$$D_{absent} = [d(f_i, f_j)]_{f_i \in A, f_j \in E}$$

The absent feature metric constraint was modeled as a contrastive loss function, which aims to maximize the distance between the absent features and the existing features. The γ is a margin hyperparameter that controls the minimum distance between the absent and existing features.

$$L_{contrastive} = \frac{1}{|A|} \sum_{i \in A} \max\left(0, \gamma - \min_j D_{absent}(i, j)\right)$$

In general, for the existing classes of elements, we can directly apply the rules of inter-class separation and intra-class aggregation. When a class is absent, we aim to separate its features as much as possible from those of the existing classes, thereby enhancing the discriminative ability for negative samples. The final feature metric loss function is:

$$L_{Metric} = L_{triplet} + L_{contrastive}$$

D. Multi-Label Classification

Given the physical attention feature $F_{PhyAtt} \in \mathbb{R}^{C \times 128}$, we predicted probabilities $\hat{P} = [P_1, P_2, \dots, P_{C-1}, P_{bg}] \in \mathbb{R}^C$ for each category in the mixed radionuclide energy spectrum using global sum pooling and the Sigmoid function.

$$S_i = \sum_{k=1}^{128} F_{PhyAtt}(i, k), \quad 1 \leq i \leq C, \quad S \in \mathbb{R}^C$$

As shown in the equation above, S follows a clear linear computation process with F_{PhyAtt} , enhancing the causal logic and interpretability of the model inference.

$$\hat{P} = \text{Sigmoid}(S)$$

Subsequently, the value of S was transformed into a probability distribution using the Sigmoid function. In other words, the predicted probability can be considered as the linear contribution based on the feature weights. The predicted probability $\hat{P}$ is a vector of length C , where each element represents the probability of the corresponding radionuclide category being present in the energy spectrum sample. The last element, P_{bg} , represents the background probability.

The predicted probability $\hat{P}$ was then used to calculate the Binary Cross-Entropy (BCE) loss for each category, which is defined as follows:

$$L_{BCE} = \sum_{j=1}^C (-P_j \log(\hat{P}_j) - (1 - P_j) \log(1 - \hat{P}_j))$$

Since BCE loss treats each category as an independent binary classification, it overlooks the fact that the probabilities of a radionuclide existence and the background are mutually exclusive. In other words, if a radionuclide exists in the energy spectrum, the background probability should be zero, and vice versa. Therefore, a cross-entropy loss function was used to strengthen the mutually exclusive relationship between the probability of radionuclide existence and the background.

To achieve this, we first identified the maximum probability of radionuclide existence from the predicted probabilities $\hat{P}$, denoted as P_{Rmax} , and the background probability P_{bg} .

$$P_{Rmax} = \max(P_1, P_2, \dots, P_{C-1})$$

$$\hat{P}_{Rmax} = \max(\hat{P}_1, \hat{P}_2, \dots, \hat{P}_{C-1})$$

$$\hat{P}_{RB} = \text{softmax}([\hat{P}_{Rmax}, \hat{P}_{bg}])$$

$$L_{RB} = -P_{Rmax} \log(\hat{P}_{RB}^0) - P_{bg} \log(\hat{P}_{RB}^1)$$

In the above equations, $\hat{P}_{RB}$ is a two-dimensional vector representing the predicted probabilities of radionuclide existence and background, respectively. The first element corresponds to the maximum radionuclide existence probability, while the second element corresponds to the background probability.

The L_{RB} loss function is designed to enhance the mutually exclusive relationship between the radionuclide existence probability and the background probability. When a radionuclide does not exist ($P_{Rmax} = 0$), the loss function will increase $\hat{P}_{Rmax} \rightarrow 0$ and decrease the background probability $\hat{P}_{bg} \rightarrow 1$. By explicitly modeling this mutually exclusive relationship, the model's ability to discriminate the probability of radionuclide existence is improved. The background probability is also effectively suppressed, enhancing the model's robustness against background noise interference. The total MLC loss function is the sum of the BCE loss and the RB loss:

$$L_{MLC} = L_{BCE} + L_{RB}$$

Data Preparation

A. Monte Carlo Simulation

To quantitatively assess the generalizability of the model, a particle transport model based on Geant4-11.1.3 [32] was developed. This allows us to accurately control multiple variables for quantitative analysis. In this study, two scintillators were considered: a 3-inch NaI(Tl) and a 1-inch CeBr3, as shown in Figure 2. The detectors consist of a 2-mm-thick stainless steel shell containing Fe, Cr, Ni, and C, a 0.3-mm-thick MgO reflective layer, and a SiPM photoelectric converter. In contrast, CeBr3 has emerged as an alternative to traditional NaI(Tl) and LaBr3:Ce crystals due to its excellent energy resolution, negligible background, and increasing application in environmental monitoring [36–38]. NaI is widely used due to its low cost and ease of processing into larger sizes. The simulated gamma-ray energy response range spans from 30 to 3000 keV, with 100 million photons simulated per calculation. The statistical uncertainty of the simulated results ranged from 0.0002 to 0.0017. The energy deposition spectrum of gamma photons in the detector was obtained by inputting the target radionuclide energy and branching ratio.

Gaussian broadening was implemented to simulate electronic noise:

$$E = E_0 + \text{FWHM}(E_0) \cdot g$$

$$\text{FWHM}(E_0) = C_1 + C_2 \sqrt{E_0} + C_3 E_0$$

Where E_0 is the energy deposition obtained from the simulation, E is the broadened energy, $g \sim N(0, 1)$ is a random variable that follows a standard normal distribution, and C_1 , C_2 , and C_3 are coefficients fitted from experimental data, as shown in Table 1.

We considered 8 artificial radioactive nuclides (^{241}Am , ^{133}Ba , ^{57}Co , ^{131}I , ^{60}Co , ^{134}Cs , ^{137}Cs , ^{152}Eu) along with the background energy spectrum. These include common emission contaminants after nuclear accidents [33–35] and include challenges such as low-energy Compton plateau interference and peak overlapping shown as Fig. 3. The background energy spectrum was measured for 3600 seconds with the detector shielded by lead bricks.

B. Data Augmentation

Data augmentation was conducted by controlling the gross count, SNR and radionuclide mixing ratio. These methods have been widely used in previous research and proven to effectively simulate the random noise in measured energy spectra [12, 15, 19]. The parameters to be adjusted and their corresponding ranges are listed in Table 1.

C. Dataset Splitting

As shown in Table 1, the dataset was categorized into 5 types. The training data set uses an NaI(Tl) detector, with a radionuclide mixing quantity $k_{train} \leq 2$. Four test data sets were considered, each corresponding to different Gaussian broadening coefficients or detector types, with the quantity of radionuclide mixing $k_{test} = 3$. This means that the radionuclide category combinations in the test set do not appear in the training set, requiring the model to infer new energy spectrum shapes and radionuclide combinations. Notably, this is more challenging than the evaluation methods used in previous studies, which typically divided the dataset proportionally. This imposes higher demands on the model's generalization ability.

D. Model Training

During training, the NaI-Train dataset was split into an 8:2 ratio for training and validation. During each data reading step in training, Gaussian random noise with a mean of 0.1 and variance of 0.04 was added to augment the training dataset. The initial learning rate was set to $1e-4$, and parameter optimization was performed using the Adam optimizer. The overall loss function during training is:

$$L_{Total} = \lambda \times L_{Metric} + \beta \times L_{MLC}$$

λ and β control the feature metric constraints and the MLC loss weight, respectively. In practice, $\lambda = 0.02$ and $\beta = 1$.

As shown in Fig. 4, the classification loss (L_{BCE} and L_{RB}) converges rapidly during the early stages of training, followed by the feature space constraint loss ($L_{triplet}$ and $L_{contrastive}$), which guides further optimization of the model parameters. As shown in Fig. 5, the model performance on the validation set reaches preliminary convergence around Epoch = 100, with the corresponding loss curve indicating that this is primarily due to the contribution of the classification loss. Subsequently, at Epochs 150 and 300, the feature space constraint loss guides additional performance improvement.

Evaluation and Analysis

A. Evaluation Metrics

The F1 score and Model parameters were used as evaluation metrics in this article. Let Y_i be the binary label, $\hat{Y}_i$ be the predicted value converted into a binary sequence using threshold $T = 0.5$.

F1 score (F1): The harmonic mean of precision and recall, offering a balanced evaluation of model performance by accounting for both false positives and false negatives.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

In the formula, precision is the ratio of correctly predicted outputs to the total predicted outputs, averaged across all sample instances.

$$\text{Precision} = \frac{\sum_i |Y_i \cap \hat{Y}_i|}{\sum_i |\hat{Y}_i|}$$

Recall measures the ratio of correctly predicted labels to the total number of annotated labels, averaged across all sample instances.

$$\text{Recall} = \frac{\sum_i |Y_i \cap \hat{Y}_i|}{\sum_i |Y_i|}$$

Micro-averaging is applied during the evaluation process. All label predictions are aggregated, and the summary metrics are then calculated.

Model parameters: The total number of adjustable parameters in the model, including weights and biases, which are optimized during the training process. Fewer parameters generally indicate lower computational resource requirements, better interpretability, and a reduced risk of overfitting.

B. Compared Methods

To evaluate the recognition performance of the proposed method, we considered 3 classic CNN models: AlexNet [39], VGG-16 [40], and ResNet-18 [41]. VGG-16 is a representative of dense and deep CNNs, while ResNet-18 represents the residual network structure. These models were originally designed for 2D CNNs in image recognition tasks but have been adapted to 1D in this paper based on their original architecture.

Ablation experiments were conducted in this study. The test model that only implements the classification loss is referred to as “Ours_{CLS}” in subsequent tests, while the proposed method is named “Ours”.

C. Quantitative Analysis

Comprehensive performance: Based on the parameters in Table 1, $1e6$ nuclide test samples and $1e5$ background samples were generated. The overall performance and parameter counts of the models are summarized in Table 2 and Table 3.

Across all test datasets, the proposed model consistently maintains SOTA performance, even when faced with challenges such as increased mixed radionuclide quantities, variations in Gaussian broadening coefficients, and differences in detector types. Notably, the model’s parameter count is only 1.58% of that of the ResNet-18-CNN model (which has the second fewest parameters). Despite this, the F1 Score consistently remains above 95%, with improvements of at least 5.054%, 7.507%, 3.786%, and 3.847% over traditional methods for each dataset.

Ablation test results show that after introducing the feature space metric constraint, the proposed model’s F1 Score increased by 2.812% and 1.82% in the NaI(Tl) detector, while in the CeBr3 detector, the F1 Score increased by 0.775% and 0.961%. These results suggest that the feature space metric constraint method is more effective in improving the performance of low-energy resolution detectors (NaI(Tl)) in spectral recognition.

Although all DL models generally exhibit high recognition precision (as consistent with previous research findings), the proposed model excels particularly in recall performance. In comparison with traditional methods, the proposed model achieves improvements of at least 8.711%, 12.282%, 6.83%, and 6.8% on different datasets, demonstrating its advantage in reducing false negatives.

In contrast, traditional methods show considerable performance fluctuations. For instance, from the NaI-1 to NaI-2 dataset, the F1 Score of VGG-16-CNN decreased by 9.001%, the Recall of ResNet-18-CNN and AlexNet-CNN dropped by 3.897% and 5.763%. Additionally, when compared to the performance on the NaI-1 and CeBr3-1 datasets, the F1 Score of AlexNet-CNN decreased by 3.318%. This phenomenon contradicts the conventional understanding of spectral analysis. Since the CeBr3 detector has superior energy resolution, an improvement in recognition accuracy would be theoretically expected.

In summary, the proposed model demonstrates not only high generalization ability and superior performance under various experimental conditions but also a significant breakthrough in model lightweighting.

Low gross count: To assess robustness under low-count conditions, as shown in Fig. 6(a), the gross count was controlled to be 5 intervals, each of which contained $1e5$ nuclide samples. An additional $1e5$ background samples were generated. The SNR was kept above 0.5 to eliminate the influence of low-SNR conditions. As shown in Fig. 6(a), the performance of the model improved with increasing gross counts. The proposed method consistently achieved SOTA performance under all conditions.

Low SNR: The SNR quantifies the prominence of characteristic peaks relative to background noise, reflecting the relative intensities of source and background radiation. The SNR was divided into 5 intervals according to Fig. 6(b), each of which contained $1e5$ nuclide samples. An additional $1e5$ background samples were generated. The gross count was controlled to $1e5$ to exclude the effect of low counts. As shown in Fig. 6(b), the proposed model outperformed others. In particular, AlexNet outperformed other traditional models with low SNR in the NaI(Tl) detector. However, its performance on the CeBr3-1 dataset is clearly worse than that of the NaI-1 dataset. This contradicts established expectations in nuclear physics, because higher energy resolution typically facilitates spectral analysis. This highlights the interpretability challenges of traditional neural networks.

D. Explainable Analysis

We visualized the F_{PhyAtt} weights to clarify the computation process behind the predicted probabilities of the model for the labels. As shown in Fig. 7 and Fig. 8, the weight visualizations correspond to the energy spectra obtained using NaI and CeBr3 detectors, respectively. The orange curve in the figure represents the prior feature peak physical information matrix, and the Attention features F_{PhyAtt} weights are overlaid on the original energy spectrum curve with different colors. The F_{PhyAtt} weights are calculated based on the physical information matrix, which is then fine-tuned using the feature space metric constraints.

Fig. 7(a) shows the energy spectrum containing three mixed radionuclides: ^{60}Co , ^{134}Cs , and ^{137}Cs , which are common contaminants released in nuclear accidents. Due to the limited energy resolution of the NaI(Tl) detector, the characteristic peaks of 604.721 keV (^{134}Cs) and 661.7 keV (^{137}Cs) overlapped, as observed. Nonetheless, the proposed model effectively applies weights to critical regions, avoiding the 661.7 keV peak when interpreting ^{134}Cs .

Fig. 7(b) shows the energy spectrum, which includes three mixed radionuclides: ^{134}Cs , ^{137}Cs , and ^{152}Eu . The gross count is lower than that in Fig. 7(a), showing larger statistical fluctuations. The proposed model still avoids the peak region at the top of 661.7 keV peak when interpreting ^{134}Cs . Despite the weak features of ^{152}Eu on a global scale, appropriate weights can still be

applied. These observations demonstrate the effectiveness of applying physical information constraints to the radionuclide characteristic peaks.

Additionally, when interpreting ^{134}Cs , the model also considers the right half of the 661.7 keV peak. When ^{134}Cs and ^{137}Cs coexist, the FWHM of the 661.7 keV peak typically increases due to peak overlap, with the right half showing a peak-valley characteristic. This indicates that F_{PhyAtt} is not confined to the characteristic peak region, but instead considers the global features of the entire spectrum. This is the result of fine-tuning the weights following the application of feature space measurement constraints. This helps summarize similar features within the same category and ensures the discriminability of features across different categories during training.

Fig. 8(a) shows the energy spectrum, which includes three mixed radionuclides: ^{131}I , ^{60}Co , and ^{137}Cs . Its gross count is lower, with greater statistical fluctuations. For ^{131}I , the model does not assign sufficient weight to highlight features in the 80.185 keV region, which completely overlaps with the scatter peak in the low-energy region, making it indistinguishable to human experts. The model also does not emphasize the 636.989 keV (^{131}I), as it overlaps with 661.7 keV (^{137}Cs) peak. This result stems from the feature space measurement constraints. The model emphasizes the right half of the 661.7 keV (^{137}Cs) region to distinguish it from 636.989 keV (^{131}I) peak.

Fig. 8(b) shows the energy spectrum, which contains three mixed radionuclides: ^{57}Co , ^{134}Cs , and ^{152}Eu . Although the characteristic peaks of ^{134}Cs and ^{152}Eu are weak on a global scale, the model still enhances features in critical regions. Additionally, for the 1173.2 keV (^{60}Co) characteristic peak, the model enhances the weight on the right half of the region. This is because it may overlap with the 1112.1 keV (^{152}Eu) characteristic peak. Correspondingly, the model enhances the weight on the left half of the region for the 1112.1 keV (^{152}Eu). Additionally, when ^{152}Eu coexists with ^{57}Co , a phenomenon similar to that observed when ^{137}Cs and ^{131}I coexist in Fig. 8(a) occurs. The model does not emphasize the 121.8 keV (^{152}Eu), as this would result in the final eigenvector being too similar to that of ^{57}Co .

The model's ability to adaptively adjust the weights of the characteristic peaks based on the feature space metric constraints is a key factor in its success. This allows the model to effectively distinguish between radionuclides with similar features, even in challenging conditions such as low gross counts and low SNR. The model's performance is not solely dependent on the physical information matrix, but also on the fine-tuning of the weights based on the feature space metric constraints. This approach enables the model to consider the full spectrum of features and generate a discriminative feature space, ultimately improving classification performance.

E. Feature Metrics Analysis

To better demonstrate the effectiveness of the measurement constraints in the feature space, we visualized the F_{PhyAtt} weights distribution in the feature space using t-distributed Stochastic Neighbor Embedding (t-SNE) [42]. Fig. 9 illustrates the feature space distributions at the start and after training, evaluated on the CeBr3-1 test set.

In the absence of training, the feature weights are primarily determined by physical information, as shown in Fig. 9(a). At this stage, the model can highlight feature differences for each radioactive nuclide to some extent but cannot distinguish the features of missing nuclides. Under the measurement constraints of the feature space, the F_{PhyAtt} weights are fine-tuned throughout the training process, as shown in Fig. 9(b). For example, the gray curve in the figure marks the characteristic distribution of ^{152}Eu . After training, the feature distribution has been significantly separated.

After training, features of similar radioactive nuclides, such as ^{57}Co , ^{131}I , and ^{134}Cs (dashed oval outline), which initially form two separate feature clusters, become more concentrated. Conversely, a distinct feature distance is maintained between the feature set of the missing nuclide and that of the present nuclide, reducing the model misclassification rate.

Conclusions

This paper presents FhyMetric-Net, a lightweight and interpretable NN model for the identification of mixed radionuclides. The model incorporates physical constraints on prior feature peaks of radionuclides and utilizes the powerful data-driven fitting capability of NNs, can automatically infer the probability of the existence of mixed radionuclides and provides weight explanations consistent with expert knowledge.

The model is lightweight, with only 0.0616 million parameters, which is significantly smaller than traditional CNN models like VGG-16-CNN (88.843 million parameters), ResNet-18-CNN (3.877 million parameters), and AlexNet-CNN (50.100 million parameters). This lightweight design allows the model to be deployed on resource-constrained devices, making it suitable for real-time applications in the nuclear industry.

The proposed model was evaluated on a comprehensive dataset that includes various radionuclide combinations, Gaussian broadening coefficients, and detector types. The results demonstrate that the model achieves state-of-the-art (SOTA) performance, with an F1 Score consistently above 95% across all test datasets.

The model's ability to adaptively adjust the weights of characteristic peaks based on feature space metric constraints allows it to effectively distinguish between radionuclides with similar features, even in challenging conditions such as low gross counts and low SNR. This adaptability is a key factor in the model's

s success, enabling it to maintain high classification performance across different experimental conditions.

The model' s explainability is enhanced through the visualization of F_{PhyAtt} weights, which provides insights into the model' s decision-making process and aligns with expert knowledge in nuclear physics. This interpretability is crucial for building trust in the model' s predictions and ensuring its reliability in practical applications. The research results will help advance mixed radionuclide automatic identification technology in the highly reliable nuclear industry.

Bibliography

- [1] J. Klusoň, Environmental monitoring and in situ gamma spectrometry. *Radiat. Phys. Chem.* 61, 209-216 (2001). [https://doi.org/10.1016/S0969-806X\(01\)00242-0](https://doi.org/10.1016/S0969-806X(01)00242-0) [2] H. Kofuji, In situ measurement of ^{134}Cs and ^{137}Cs in seabed using underwater γ -spectrometry systems: application in surveys following the Fukushima Dai-ichi Nuclear Power Plant accident. *J. Radioanal. Nucl. Chem.* 303, 1575-1579 (2015). <https://doi.org/10.1007/s10967-014-3702-0> [3] D. Connor, P.G. Martin, T.B. Scott, Airborne radiation mapping: overview and application of current and future aerial systems. *Int. J. Remote Sens.* 37, 5953-5987 (2016). doi: 10.1080/01431161.2016.1252474 [4] E.G. Androulakaki, M. Kokkoris, C. Tsabaris, et al., In situ spectrometry in a marine environment using full spectrum analysis for natural radionuclides. *Appl. Radiat. Isot.* 114, 76-86 (2016). <https://doi.org/10.1016/j.apradiso.2016.05.008> [5] D. Fagan, S. Robinson, R. Runkle, Statistical methods applied to gamma-ray spectroscopy algorithms in nuclear security missions. *Appl. Radiat. Isot.* 70, 2428-2439 (2012). <https://doi.org/10.1016/j.apradiso.2012.06.016> [6] J. Routti, S. Prussin, Photopeak method for the computer analysis of γ -ray spectra from semiconductor detectors. *Nucl. Instrum. Methods.* 72, 125-142 (1969). [https://doi.org/10.1016/0029-554X\(69\)90148-7](https://doi.org/10.1016/0029-554X(69)90148-7) [7] I.A. Slavić, S.P. Bingulac, A simple method for full automatic gamma-ray spectra analysis. *Nucl. Instrum. Methods.* 84, 261-268 (1970). [https://doi.org/10.1016/0029-554X\(70\)90270-3](https://doi.org/10.1016/0029-554X(70)90270-3) [8] G. Xiao, L. Deng, B. Zhang et al., A nonlinear wavelet method for low-level gamma-ray spectra smoothing. *J. Nucl. Sci. Technol* 41, 73-76 (2004). <https://doi.org/10.3327/jnst.41.73> [9] C.J. Sullivan, S.E. Garner, K.B. Butterfield, Wavelet analysis of gamma-ray spectra. *IEEE Symposium Conference Record Nuclear Science 2004.* 1 pp. 281-286 Vol. 1 (2004). <https://doi.org/10.1109/NSSMIC.2004.1462198> [10] C.J. Sullivan, M.E. Martinez, S.E. Garner, Wavelet analysis of sodium iodide spectra. *IEEE Nuclear Science Symposium Conference Record, 2005* 1 pp. 302-306 (2005). <https://doi.org/10.1109/NSSMIC.2005.1596258> [11] M. Monterial, K. Nelson, S. Labov, et al., Benchmarking Algorithm for Radio Nuclide Identification (BARNI) Literature Review. (2019,2). <https://www.osti.gov/biblio/1544518> [12] S. Qi, W. Zhao, Y. Chen et al., Comparison of machine learning approaches for radioisotope identification using NaI(Tl) gamma-ray spectrum. *Appl. Radiat. Isot.* 186 pp. 110212 (2022).

<https://doi.org/10.1016/j.apradiso.2022.110212> [13] S.M. Galib, P.K. Bhowmik, A.V. Avachat, et al., A comparative study of machine learning methods for automated identification of radioisotopes using NaI gamma-ray spectra. Nucl. Eng. Technol. 53, 4072-4079 (2021). <https://doi.org/10.1016/j.net.2021.06.020> [14] C. Li, S. Liu, C. Wang, et al., A new radionuclide identification method for low-count energy spectra with multiple radionuclides. Appl. Radiat. Isotopes. 185, 110219 (2022). <https://doi.org/10.1016/j.apradiso.2022.110219> [15] S. Qi, S. Wang, Y. Chen, et al., Radionuclide identification method for NaI low-count gamma-ray spectra using artificial neural network. Nucl. Eng. Technol. 54, 269-274 (2022). <https://doi.org/10.1016/j.net.2021.07.025> [16] Liu, HL., Ji, HB., Zhang, JM, et al., Identification algorithm of low-count energy spectra under short-duration measurement based on heterogeneous sample transfer. NUCL SCI TECH. 54, 42 (2025). <https://doi.org/10.1007/s41365-024-01595-y> [17] Sarker, I.H, Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. SN COMPUT. SCI. 2, 420 (2021). [10.1007/s42979-021-00854-1](https://doi.org/10.1007/s42979-021-00854-1) [18] Mario Gomez-Fernandez, Weng-Keen Wong, Akira Tokuhito, et al., Isotope identification using deep learning: An explanation. Nucl. Instrum. Meth. A 988, 164925 (2021). [10.1016/j.nima.2020.164925](https://doi.org/10.1016/j.nima.2020.164925) [19] Yu Wang and Qingxu Yao and Qianhu Zhang, et al., Explainable radionuclide identification algorithm based on the convolutional neural network and class activation mapping. Nucl. Eng. Technol. 54, 4684-4692 (2022). [10.1016/j.net.2022.08.011](https://doi.org/10.1016/j.net.2022.08.011) [20] Zakariya Chaouai, Geoffrey Daniel, Jean-Marc Martinez, et al., Application of adversarial learning for gamma-ray radionuclides identification spectra. Nucl. Instrum. Meth. A. 1033, 166670 (2022). <https://doi.org/10.1016/j.nima.2022.166670> [21] Liu, HL., Ji, HB., Zhang, JM., et al. A novel approach for feature extraction from a gamma-ray energy spectrum based on image descriptor transferring for radionuclide identification. NUCL SCI TECH. 33, 158 (2022). <https://doi.org/10.1007/s41365-022-01150-7> [22] Liu, HL., Ji, HB., Zhang, JM., et al. Novel algorithm for detection and identification of radioactive materials in an urban environment. NUCL SCI TECH. 34, 154 (2023). <https://doi.org/10.1007/s41365-023-01304-1> [23] Vaswani, Ashish, Parmar N, et al. Attention is all you need. in Paper In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS' 17) (Red Hook, USA, 2017). <https://doi.org/10.5555/3295222.3295349> [24] Niu, Zhaoyang., Zhong, Guoqiang., Yu, Hui., A review on attention mechanism of NEUROCOMPUTING. (2021). <https://doi.org/10.1016/j.neucom.2021.03.091> [25] Zhu, Feng., Li, Hongsheng., Ouyang, Wanli., et al. Learning Spatial Regularization With Image-Level Supervisions for Multi-Label Image Classification. in Paper Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (Hawaii, USA, 2017) <https://doi.org/10.1109/CVPR.2017.219> [26] Gao, Bin-Bin and Zhou, Hong-Yu, et al. Learning to Discover Multi-Class Attentional Regions for Multi-Label Image Recognition, IEEE Transactions on Image Processing 30, 5920-5932 (2021). <https://doi.org/10.1109/TIP.2021.3088604.721> [27] Jiaqian Sun, De-qing Niu, Jie Liang, et al. Rapid nuclide identification algorithm based on

self-attention mechanism neural network. *Ann. Nucl. Energy.* 207, 110708 (2024). <https://doi.org/10.1016/j.anucene.2024.110708> [28] Wang Y, Zhang Q, Yao Q, et al. Multiple radionuclide identification using deep learning with channel attention module and visual explanation. *Front. Phys.* 10 1036557 (2022). doi: 10.3389/fphy.2022.1036557 [29] Schlagenhauf, T., Lin, Y. Noack, B. Discriminative feature learning through feature distance loss. *Machine Vision and Applications* 34, 25 (2023). <https://doi.org/10.1007/s00138-023-01442-5> [30] KAYA, M.; BİLGE, H. Ş. Deep Metric Learning: A Survey. *Symmetry.* (2019). <https://doi.org/10.3390/sym11081066> [31] J. Lu, J. Hu and J. Zhou, Deep Metric Learning for Visual Understanding: An Overview of Recent Advances, *IEEE Signal Processing Magazine*, 34, 76-84 (2017). <https://doi.org/10.1109/MSP.2017.2732900> [32] S. Agostinelli, J. Allison, K. Amako, et al., Geant4—a simulation toolkit. *Nucl. Instrum. Methods Phys. Res. A.* 506, 250-303 (2003). [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8) [33] J. Li, S. Liu, Y. Zhang et al., Pre-assessment of dose rates of ^{134}Cs , ^{137}Cs , ^{60}Co for marine biota from discharge of Haiyang Nuclear Power Plant, *J. Environ. Radioactiv.* 147 pp. 8-13 (2015). <https://doi.org/10.1016/j.jenvrad.2015.05.001> [34] S. Ueda, H. Hasegawa, H. Kakiuchi et al., Fluvial discharges of radio-caesium from watersheds contaminated by the Fukushima Dai-ichi Nuclear Power Plant accident, Japan. *J. Environ. Radioactiv.* 118 pp. 96-104 (2013). <https://doi.org/10.1016/j.jenvrad.2012.11.009> [35] G. Katata, M. Ota, H. Terada et al., Atmospheric discharge and dispersion of radionuclides during the Fukushima Dai-ichi Nuclear Power Plant accident. Part I: Source term estimation and local-scale atmospheric dispersion in early phase of the accident. *J. Environ. Radioactiv.* 109 pp. 103-113 (2012). <https://doi.org/10.1016/j.jenvrad.2012.02.006> [36] Christos Tsabaris, Efrossyni G. Androulakaki, Aristides Prospathopoulos, et al., Development and optimization of an underwater in-situ cerium bromide spectrometer for radioactivity measurements in the aquatic environment. *J. ENVIRON. RADIOACTIV.* 204, 12-20 (2019). <https://doi.org/10.1016/j.jenvrad.2019.03.021> [37] Wang, M., Gu, Y., Xiong, ML, et al. Method for rapid warning and activity concentration estimates in online water γ -spectrometry systems. *NUCL SCI TECH.* 35, 49 (2024). <https://doi.org/10.1007/s41365-024-01395-4> [38] Gu, Y., Sun, K., Ge, LQ, et al. Investigating the minimum detectable activity concentration and contributing factors in airborne gamma-ray spectrometry. *NUCL SCI TECH.* 32, 110 (2021). <https://doi.org/10.1007/s41365-021-00951-6> [39] Krizhevsky, Alex, Sutskever, Ilya, Hinton, Geoffrey E, et al. ImageNet Classification with Deep Convolutional Neural Networks. in Paper Presented at the Advances in Neural Information Processing Systems (NIPS) (Lake Tahoe, USA, 2012). <https://doi.org/10.1145/3065386> [40] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in Paper Presented at the 3rd International Conference on Learning Representations (San Diego, USA, 2015) <https://doi.org/10.48550/arXiv.1409.1556> [41] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in Paper Presented at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Las Vegas, USA, 2016) <https://doi.org/10.1109/CVPR.2016.90> [42] L. Van der

Maaten, G. Hinton, Visualizing data using t-SNE. J. Mach. Learn. Res. 9(11), 2579-2604.721 (2008)

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv –Machine translation. Verify with original.