



YOLOSpecNN: a novel γ -ray spectra full-energy peak automatic search and segmentation model inspired by YOLO

Cao-Lin Zhang^{1,2} · Jiang-Mei Zhang^{1,2} · Hao-Lin Liu^{1,2,3} · Shu-Ya Qin⁴ · Jia-Qi Wang^{1,2}

Received: 10 November 2024 / Revised: 5 January 2025 / Accepted: 7 January 2025 / Published online: 2 May 2026

© The Author(s), under exclusive licence to China Science Publishing & Media Ltd. (Science Press), Shanghai Institute of Applied Physics, the Chinese Academy of Sciences, Chinese Nuclear Society 2026

Abstract

Qualitative identification and analysis of radioactive nuclides in unknown environments are essential for the remote monitoring and prompt early warning of radioactive contamination. In recent years, deep learning techniques have made significant strides in automated qualitative identification. However, the quantitative analysis of radioactive nuclides still depends on traditional methods to determine peak positions and boundaries. These methods often require extensive manual expertise and parameter tuning and thus fail to meet the demands of unmanned remote monitoring. This paper presents a novel framework for automatic full-energy peak segmentation, called YOLOSpecNN. We introduce a multi-root mean square error joint optimization function and a unified regression model capable of simultaneously predicting the central position, boundaries, and confidence of full-energy peaks. To address the challenge of low recall rates due to narrow, low-intensity, and overlapping peaks, we propose a new multiscale context feature extraction module (MSNN module). This module effectively enhances the local detailed features and significantly improves the recall rates. The effectiveness of the proposed method is validated using six artificial radioactive nuclides (^{241}Am , ^{57}Co , ^{131}I , ^{134}Cs , ^{137}Cs , and ^{60}Co) along with ^{40}K , and a mixed-energy spectrum dataset is constructed for quantitative evaluation. The experimental results show that the proposed method significantly outperforms traditional approaches, achieving a precision of 0.998, a recall of 0.95, the best F1 score of 0.974@0.427, and an average precision of 0.946. Compared with traditional morphological methods, the proposed method improves the precision, recall, and best F1 score by 0.512, 0.199, and 0.391, respectively. Ablation experiments further reveal that the MSNN module notably enhances the recall by 0.067. Moreover, the proposed method performs excellently even in challenging environments with low gross counts and a low signal-to-noise ratio (SNR), achieving state-of-the-art results. Additionally, the model achieves an average real-time inference performance of 16.1941 ms on a device with a 15-W low-power budget. Overall, the proposed method demonstrates exceptional performance in the automatic search and segmentation of full-energy peaks, offering robust support for the implementation of unmanned remote radiation monitoring systems.

Keywords Gamma spectroscopy · Gamma ray spectral analysis · Peak searching and segmentation · Nuclide identification

This work was supported by the Sichuan Province “Open competition to select the best candidate” Project (No. 24zs9102).

✉ Jiang-Mei Zhang
zjm@swust.edu.cn

¹ School of Information Engineering, Southwest University of Science and Technology, Mianyang 621010, China

² CAEA Innovation Center of Nuclear Environmental Safety Technology, Southwest University of Science and Technology, Mianyang 621010, China

³ Department of Automation, University of Science and Technology of China, Hefei 230026, China

⁴ Oxford Brookes College, Chengdu University of Technology, Chengdu 610059, China

1 Introduction

Gamma spectroscopy is crucial for analyzing the energy peaks and intensities of gamma rays, identifying the characteristic energies of different radionuclides, and determining their types and activities. These provide valuable data for nuclear physics research [1], environmental monitoring [2, 3], medical applications [4], and homeland security [5].

In gamma spectroscopy analysis, precisely defining the area of each full-energy peak in the γ -ray spectra is essential for accurately determining the isotope energy positions and reducing background noise interference. Clear peak

boundaries optimize the integration area, aiding in the calculation and quantitative analysis of the peak areas. This enables researchers to calculate the radioisotope radioactivity and assess the safety and environmental impact of the samples.

However, challenges arise due to statistical fluctuations from low-count spectra's discrete Poisson fluctuations, low signal-to-noise ratio (SNR), and peak overlap caused by limited energy resolution. These issues complicate automatic peak search and region segmentation [6–8].

The gamma ray spectrum analysis of radioisotopes generally involves two primary steps: peak search and peak area fitting [9, 10]. Early peak search methods required smoothing to reduce statistical fluctuations [11], followed by the development of the Fourier transform and wavelet-based techniques [12–16]. Recently, peak search methods based on image morphology have shown superior performance for complex full-energy peaks compared with traditional methods [17, 18]. However, factors such as the size of the morphological structural element, smoothing parameters, and false-peak screening thresholds require manual adjustments based on expert knowledge.

These methods rely heavily on human expertise to optimize parameter adjustments and balance noise smoothing, while retaining peak features. This reliance limits the automation of radioisotope identification devices, rendering them unsuitable for unmanned remote monitoring in communication-limited or harsh environments.

With the rapid development of deep learning (DL) technology, researchers have increasingly explored its potential for automatic feature extraction and radioisotope identification [19–21]. DL models have demonstrated high accuracies in recognizing radionuclides [22, 23]. However, end-to-end multilabel classification models typically provide only qualitative probabilities for nuclide presence without directly outputting the location and boundaries of the characteristic peaks, which are often necessary for accurate quantitative analysis.

Recent studies also explored neural networks for spectral analysis and peak detection [24, 25]. These methods directly output the locations of characteristic peaks, offering a promising alternative to traditional radionuclide classification. However, accurate and automatic segmentation of full-energy peak regions remains unresolved, necessitating further fitting to determine the optimal peak boundaries. Challenges such as peak stacking and shape distortion mean that parameters such as the fitting areas and peak shape functions still require expert intervention.

Fortunately, the advancements in computer vision technology [26] are similar to those of our task, particularly in object detection and instance segmentation [27, 28]. These tasks aimed to distinguish the foreground from the background, predict object regions, and identify individuals in the same category. Similarly, the automatic separation of full-energy

peaks from the background can be achieved. In computer vision, object detection typically involves predicting bounding boxes (x_1, x_2, y_1, y_2) and object category probabilities, effectively answering the question: “Where are the objects?” [27]. In this study, we aimed to answer the question, “Where are the peaks?” The goal is to segment the peak regions by learning two boundaries (x_1, x_2) along the energy axis from a 1D spectrum and to distinguish the peaks at various positions.

Inspired by the You Only Look Once (YOLO) object detection algorithm [29], we proposed a novel full-energy peak region segmentation model called YOLOSpcNN. This model predicts the center, width, and confidence of the full-energy peaks in the energy spectrum, and further calculates the left and right boundaries. To address challenges such as low-intensity, narrow, and overlapping peaks, we introduced a multiscale context feature extraction module that enhances local details and improves recall performance.

Using Monte Carlo simulations, we created an NaI(Tl) detector dataset for quantitative evaluation, incorporating factors such as gross counts, SNR, and nuclides mixing. The experimental results demonstrated the superior performance of the proposed method.

2 Materials and methods

2.1 Backbone network

This study adapted the ResNet-18 architecture [30] for a one-dimensional implementation. The backbone network consists of repeated residual blocks of varying output sizes [31]. In particular, the proposed block structure incorporates skip connections, allowing the network to learn the residual mapping between the input and output. This effectively mitigates issues such as vanishing and exploding gradients in deep neural network training.

This study retains only the convolutional layers from the original ResNet-18 model and replaces the fully connected layers with a multiscale CNN (MSNN) module. The model input is a one-dimensional vector representing the energy spectrum ($X \in \mathbb{R}^{1 \times 1024}$). The input energy spectrum is processed using a convolutional neural network to obtain the initial feature map (F_{resnet}) and size $(1 \times 64 \times 512)$.

$$F_{\text{resnet}} = f_{\text{resnet}}(x; \theta_{\text{resnet}}), F_{\text{resnet}} \in \mathbb{R}^{(1 \times 64 \times 512)} \quad (1)$$

2.2 MSNN module

In the radionuclide spectra, some full-energy peaks covered large areas. Larger convolution kernels were required

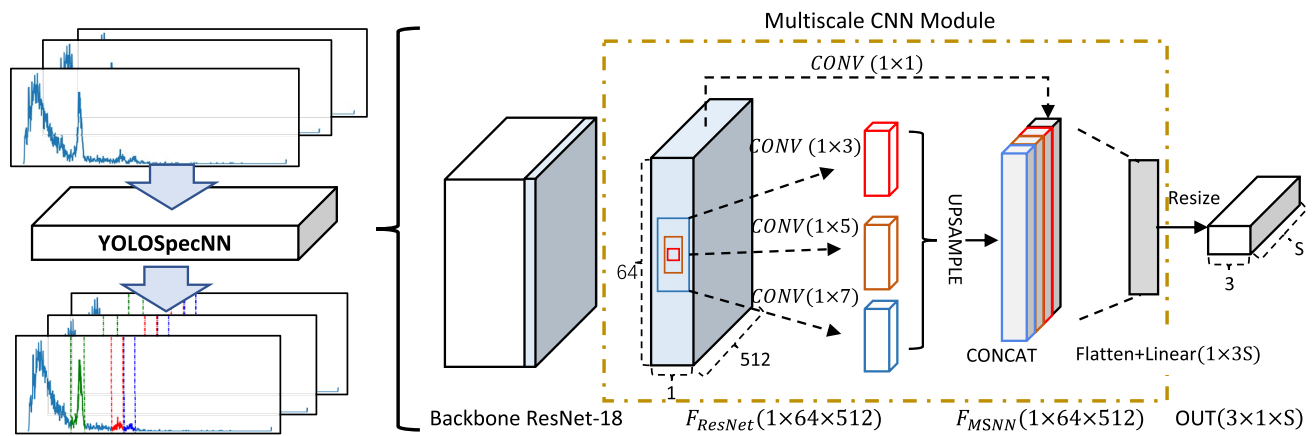


Fig. 1 (Color online) Overall structural flowchart

to capture larger spatial features. However, some radionuclide peaks are small (e.g., the full-energy peak of ^{241}Am occupies approximately $3/1024$ of the spectrum), or mixed radionuclide spectra contain overlapping peaks with similar positions (e.g., 662 keV for ^{137}Cs and 605 keV for ^{134}Cs). In such cases, smaller convolution kernels are necessary to capture finer features.

The input spectrum is processed through the backbone network to extract a deep and dense feature map using a (1×3) convolution kernel, similar to the original ResNet-18 architecture. This study introduced an MSNN module for multiscale convolutional feature fusion designed to model the full-energy peak context of the spectrum, as shown in Fig. 1. By aggregating the multiscale context across different spectral regions, the model can simultaneously capture both local and global contextual information. This concept is widely used in computer vision. Contextual features refer to the spatial arrangement and relative positioning of pixels, which can enhance the performance in areas such as small object detection [32] and visual place recognition [33, 34].

The MSNN module uses four convolution kernels of different sizes: (1×1) , (1×3) , (1×5) and (1×7) . These convolution kernels extract features from the input spectrum, generating feature maps of varying sizes, each with 128 channels. To align and aggregate feature maps at multiple scales, upsampling using nearest-neighbor interpolation is applied, resizing them to $(1 \times 64 \times 128)$. The feature maps are then concatenated, resulting in a multiscale context-aggregated feature map (F_{MSNN}) with a size of $(1 \times 64 \times 512)$.

$$F_{\text{MSNN}} = f_{\text{MSNN}}(x; \theta_{\text{MSNN}}), F_{\text{MSNN}} \in \mathbb{R}^{(1 \times 64 \times 512)} \quad (2)$$

The final multiscale feature map is flattened into a one-dimensional vector, followed by incorporating a fully connected layer. Finally, the linear output vector is resized to

$(3 \times 1 \times S)$. If the center of the full-energy peak falls within a grid cell, then the cell is responsible for detecting the object. Thus, the proposed model divides the input energy spectrum into $(1 \times S)$ grids.

$$S_{\min} = \text{ceil}(1024/S) \quad (3)$$

As shown in Fig. 2, each grid cell predicts one bounding box and its associated confidence scores, specifically, the center's coordinates (x_c), width (w), and confidence score (c).

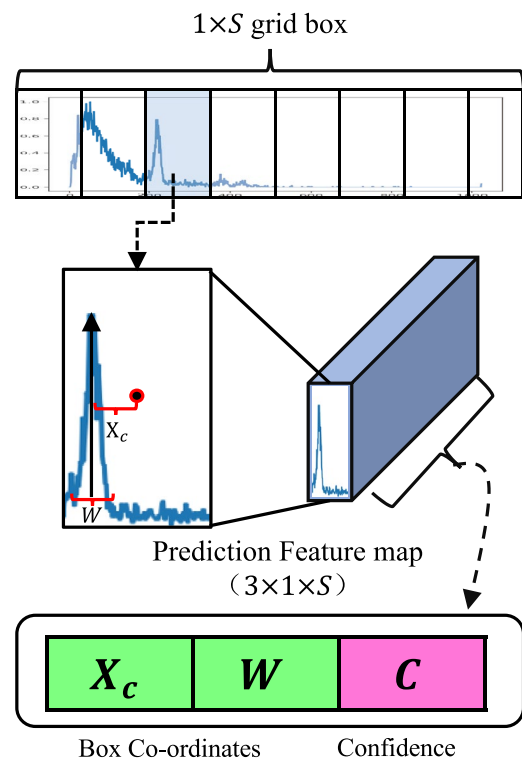


Fig. 2 (Color online) Feature map predicted by the model

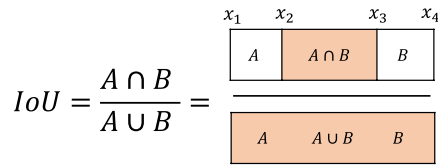


Fig. 3 (Color online) IoU calculation schematic diagram

The coordinate (x_c) represents the center of the bounding box relative to the grid cell boundaries. The width (w) is predicted relative to the entire spectrum. The confidence score (c) reflects the model's confidence in the presence of a peak within the box. If no peak is predicted in the cell, then the confidence score is zero. The final confidence prediction represents the Intersection over Union (IoU) between the predicted bounding box and the ground truth box.

The IoU measures the overlap between bounding boxes and is calculated as the ratio of the intersection area to the union area of the two regions. In this study, the boundaries are one-dimensional coordinates, which differ from those in image-based domains. Therefore, the “area” in this context is expressed as the difference in coordinates, as shown in Fig. 3. Specifically, A and B are predicted boxes with coordinate ranges $[x_1, x_3]$ and $[x_2, x_4]$, respectively. The IoU is calculated as follows:

$$\begin{cases} A \cap B = x_3 - x_2 \\ A \cup B = x_4 - x_1 \end{cases} \quad (4)$$

All the output variables are constrained within the range $[0 - 1]$, which is achieved using the sigmoid function. Its mathematical representation is expressed as:

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (5)$$

The final output of the network is a $(3 \times 1 \times S)$ tensor of the predictions. In this study, $S = 384$, and the minimum resolution for the full-energy peak center interval is three channel bins.

2.3 Loss function

During network training, three loss functions are jointly optimized for backpropagating the target boundary and confidence parameters, as shown in Eq. (6).

$$\mathcal{L}_{\text{total}}(x_i, w_i, C_i) = \mathcal{L}_{\text{objloc}}(x_i, w_i) + \mathcal{L}_{\text{objcon}}(C_i) + \mathcal{L}_{\text{noobjcon}}(C_i) \quad (6)$$

The loss function expressions for the three components are provided in Eqs. (7)–(9) as follows:

$$\begin{aligned} \mathcal{L}_{\text{objloc}}(x_i, w_i) = & \lambda_{\text{coord}} \sum_{i=0}^S i_i^{\text{obj}} (x_i - \hat{x}_i)^2 \\ & + \lambda_{\text{coord}} \sum_{i=0}^S i_i^{\text{obj}} (\sqrt{w_i} - \sqrt{\hat{w}_i})^2 \end{aligned} \quad (7)$$

$$\mathcal{L}_{\text{objcon}}(C_i) = \lambda_{\text{coord}} \sum_{i=0}^S i_i^{\text{obj}} (C_i - \hat{C}_i)^2 \quad (8)$$

$$\mathcal{L}_{\text{noobjcon}}(C_i) = \lambda_{\text{noobj}} \sum_{i=0}^S i_i^{\text{noobj}} (C_i - \hat{C}_i)^2 \quad (9)$$

$\mathcal{L}_{\text{objloc}}(x_i, w_i)$ optimizes the predicted peak center coordinates (x_i) and width (w_i) to minimize the error between the predicted and labeled values ($\hat{x}_i$) and ($\hat{w}_i$). In addition, to reduce the difference between large and small areas, we take the square root of the width (w).

$\mathcal{L}_{\text{objcon}}(C_i)$ minimizes the error between the predicted confidence (C_i) and the labeled confidence ($\hat{C}_i$) when an object is present ($i_i^{\text{obj}} = 1$) for the corresponding $\text{bbox}_i, i \in [0, S]$.

$\mathcal{L}_{\text{noobjcon}}(C_i)$ optimizes confidence when no object is present ($i_i^{\text{noobj}} = 1$) in bbox_i .

We optimized the sum of the squared errors in the model output. The corresponding weights are used to adjust the contributions of positioning errors and confidence in nontarget regions to the overall network error, thus preventing the gradient from being dominated by large noise baseline areas.

In this study, we increase the loss of the bounding box coordinate prediction and reduce the loss of confidence prediction when no object is present using two parameters, λ_{coord} and λ_{noobj} , with $\lambda_{\text{coord}} = 1$ and $\lambda_{\text{noobj}} = 0.5$.

3 Data preparation

Various random factors such as background conditions, measurement time, source intensity, and nuclide mixing can affect the spectra of radionuclides measured in real-world environments. However, acquiring large-scale measured datasets is challenging. We focused on three controllable variables—SNR, gross count, and nuclide mixing—to quantitatively evaluate our model. Monte Carlo simulations were conducted and data augmentation was applied to address these variables.

3.1 NaI(Tl) detector

NaI(Tl) detectors are widely used in gamma spectroscopy because of their high light yield, low background, low cost, ability to operate at room temperature, and ease of manufacturing into large sizes. However, their relatively low energy resolution can cause peak overlap, making automatic spectral analysis challenging.

In this study, a homemade NaI(Tl) detector was modeled using Monte Carlo simulations, as shown in Fig. 4. The detector consisted of a 2-mm-thick stainless-steel shell containing Fe, Cr, Ni, and C; a 3-inch cylindrical NaI(Tl) scintillator; a 0.3-mm-thick MgO reflective layer; and a SiPM photoelectric converter.

3.2 Monte Carlo simulation

The simulated spectra of the six common artificial radionuclides (listed in Table 1) at various distances were generated using Geant4–11.1.3 [35]. The simulation model included a point source with a gamma ray energy range of 40–3000 keV. Each simulation involved calculation of 100 million photons. We used FTFP_BERT as the reference physics list. G4EmStandardPhysics was added to the physics list to simulate the physical processes of the photoelectric reaction. The energy deposition spectrum of the detector was obtained by inputting the nuclide energy and branching ratio. The relative statistical uncertainty of the simulated results ranged from 0.02% to 0.16%. Gaussian broadening was applied to simulate the electronic noise.

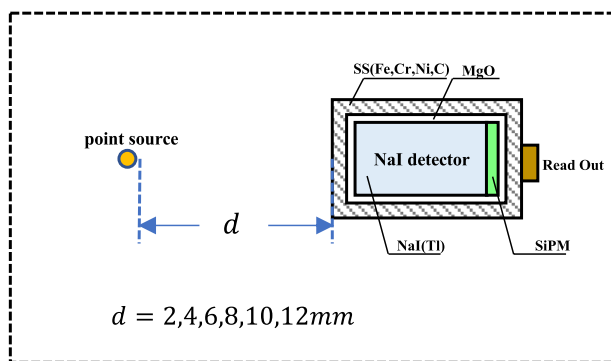


Fig. 4 (Color online) Gamma detector structure and Monte Carlo simulation layout

Table 1 Data augmentation parameter table

Radioisotope	SNR	Gross counts	Mixed quantity	Initial mixed ratio
^{241}Am , ^{57}Co , ^{131}I , ^{134}Cs , ^{137}Cs , ^{60}Co , ^{40}K	0.1~1	$1 \times 10^3 \sim 1 \times 10^5$	1~5 Artificial Nuclides + ^{40}K	0.1~0.5

$$FWHM(E_0) = C_1 + C_2 \sqrt{E_0} + C_3 E_0^2 \quad (10)$$

$$E = \frac{FWHM}{2\sqrt{2\ln 2}}g + E_0 \quad (11)$$

where E_0 is the energy deposition from the simulation; E is the broadened energy; $g \sim N(0, 1)$ is a random variable following the standard normal distribution; and C_1 , C_2 , and C_3 are the experimental coefficients with values of -0.0067 , 0.0612 , and -0.0451 , respectively.

3.3 Data augmentation

Data augmentation was performed using random gross counts, SNR, and nuclide mixing ratios. These methods have been widely used in previous studies and have proven to be effective in simulating statistical fluctuations and peak overlap in measured energy spectra [8, 22, 24].

Random gross count and random SNR: random gross count (C_{gross}) is achieved by sampling the spectrum as a probability distribution across energy.

$$C_{\text{gross}} = C_{\text{source}} + C_{\text{bq}} \quad (12)$$

$$SNR = C_{\text{source}} / (C_{\text{source}} + C_{\text{bq}}) \quad (13)$$

Given a randomly selected low SNR, we calculated the background count (C_{bq}) and nuclide count (C_{source}) by sampling. A background spectrum was obtained by measuring the shielded detector with lead bricks for 3600 s, which included natural radiation from the ^{40}K isotope. Then, the simulated spectrum is linearly combined with the measured background spectrum under different SNR conditions.

Nuclide mixing and superposition: the count of each nuclide was obtained separately based on gross count, SNR, and random proportion and then linearly superimposed. Mixed nuclides are generated through permutations and combinations. Assuming n mixed nuclides, n ratios k_n are randomly generated. To prevent large variations in nuclide counts, the initial mixed-ratio range was set between 0.1 and 0.5.

$$K = \{k_1, k_2, k_3 \dots k_n\}, k_n \sim \text{unif}([0.1, 0.5]) \quad (14)$$

Then, we normalize K values to the sum 1:

$$K_{\text{norm}} = K / \text{sum}(K) \quad (15)$$

According to the gross count C_{source} required by the random SNR, the count C_s for each nuclide source is obtained as follows:

$$C_s = \{c_{s1}, c_{s2}, c_{s3} \dots c_{sn}\} = C_{\text{source}} \times K_{\text{norm}} \quad (16)$$

Data augmentation can be performed based on the number of samples required to meet the requirements of training and performance evaluation.

Figure 5 shows the comparison between the simulated synthetic spectrum (orange) and the measured spectrum (blue). For the simulated spectra, we manually adjusted the counts of ^{137}Cs and the background to 3×10^3 , and the counts of ^{60}Co to 1×10^3 . We used a detector with the same structure as shown in Fig. 4 and obtained the measured energy spectra based on ^{137}Cs and ^{60}Co point sources. We calculated the cosine similarity between the simulated synthetic spectrum and measured spectrum as 0.965, indicating a good fit between the two.

The constructed dataset generally included six common artificial radionuclides and a natural radionuclide, ^{40}K . Among them, ^{241}Am and ^{57}Co have low energies, and their full-energy peaks occupy a small area that overlaps the strong low-energy Compton scattering region. ^{131}I , ^{60}Co , ^{134}Cs , and ^{137}Cs are commonly released during nuclear power plant accidents [36–38]. When they coexist with the natural radionuclide background ^{40}K , full-energy peak overlap may occur.

For each nuclide category, the central position (x_c), area width (w), and corresponding confidence (c) of the full-energy peaks are labeled in all the spectral samples to construct the dataset used in this study. All the energy spectra

are normalized to ensure that the input depends solely on the shape.

4 Evaluation

4.1 Evaluation metrics

Evaluation metrics, widely used in computer vision object detection models, including precision (P), recall (R), F1 score (F1), and average precision (AP), were used to evaluate our method [39, 40].

Recall: the percentage of correctly detected targets relative to the total number of targets.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (17)$$

Precision: the proportion of correctly detected targets relative to the total predicted targets.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (18)$$

F1 score: the harmonic mean of precision and recall, which evaluates model performance in a balanced manner, considering both false positives and false negatives.

$$F1 = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (19)$$

where TP denotes the number of correctly identified targets, FP denotes the number of falsely detected targets, and FN denotes the number of undetected targets.

AP: a measure of the area under the precision–recall (P–R) curve, with recall on the x -axis and precision on the y -axis. AP effectively evaluates the average detection precision across varying recall levels. The calculation formula for AP is shown in Eq. (20).

$$\text{AP} = \int_0^1 P_i(R_i) dR_i \quad (20)$$

4.2 Compared methods

Morphological-based: Luo et al. [17] compared five traditional peak search methods with a new morphology-based method using mixed spectra generated from multiple nuclides. These results indicate that the morphological method is more suitable for peak searches in the mixed spectra generated from multiple nuclides. Therefore, this method served as a baseline for comparison in this study.

The morphological peak search method consists of three steps: spectrum smoothing, morphological top-hat transformation, and false-peak screening. Finally, the significant

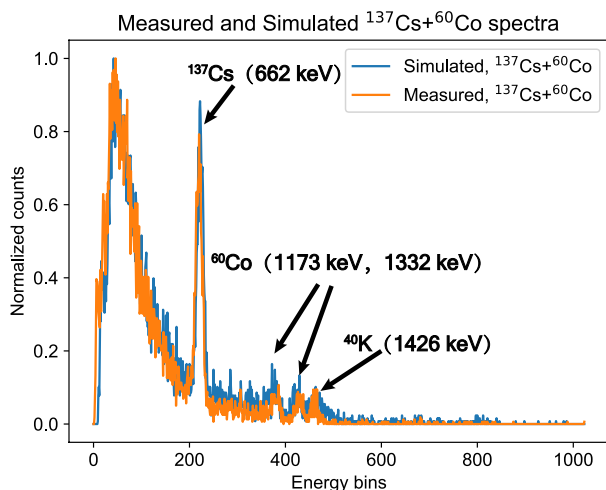


Fig. 5 (Color online) Measured and simulated $^{137}\text{Cs}+^{60}\text{Co}$ mixed spectra

peaks in the energy spectra are retained. Spectral smoothing is performed using the Savitzky–Golay filter [41]. The parameters to be adjusted are the window size (w) and polynomial order (k). As w increases and k decreases, the smoothing effect becomes more pronounced. In the morphological top-hat transformation, the structural element size (L) should be adjusted. If L is extremely small, then excessive noise is detected. If L is extremely large, then the ability of the algorithm to distinguish between adjacent peaks diminishes. After morphological white-hat transformation, the spectrum is divided into continuous nonzero regions.

The width (w_i) of each region is calculated, and the threshold (W_T) is set. If $w_i < W_T$, then it is classified as a false peak. Multiple candidate peaks are detected. When the difference between the detected peak and ground truth is within the energy window (set to 2% in this study), the search is deemed correct, whereas the others are classified as false peaks [42].

The four parameters were manually adjusted. The adjusted parameters and their ranges are presented in Table 2. The best parameter configuration was identified by iterating different parameters and using F1 as the evaluation metric, as listed in Table 2. In subsequent tests, it was labeled *MorPh*.

ResNet-18: an ablation test was performed to evaluate the effectiveness of the proposed MSNN module. To ensure experimental fairness, the MSNN module was removed from the architecture (Fig. 1). The feature map F_{resnet} was then directly expanded and connected to the following fully connected layer, labeled *ResNet-18*, in the test. The method proposed in this study was labeled *ResNet-18+MSNN*.

4.3 Model training

During the model training process, the dataset was split into an 80% training set and 20% validation set. The initial learning rate was set to $lr = 1 \times 10^{-4}$, and the batch size was set to 256. The learning rate decreased by a factor of 10 every 30 epochs. The Adam optimizer was used for parameter optimization.

Notably, during each data reading in the training process, uniformly distributed noise in the range of 0.001~0.02 and peak displacements in the range of 0~200 were randomly

added to further augment the dataset and ensure its diversity. These two data augmentation steps were applied only during training to ensure that the test dataset remained exclusive and did not overlap with the training data.

Figure 6a presents the trend of the total loss curves during the training process. Figure 6b shows the performance of various indicators on the validation dataset, with an upward trend corresponding to the loss.

4.4 Evaluation results and analysis

A comprehensive performance evaluation was first performed, followed by control of the gross count, SNR, and nuclide category for quantitative analysis. Subsequently, qualitative evaluation and inference time performance were assessed.

Comprehensive performance: values for the SNR and gross count were randomly selected based on the range specified in Table 1, and a test set of 60,000 samples is generated. Figure 7 and Table 3 show the recall and precision performances of the three methods for the test set.

The DL-based method calculates the corresponding recall and precision at different confidence thresholds and generates a P–R curve. The morphology-based method visualizes the calculation results during the parameter iteration process in a scatter plot. The DL-based method outperformed the morphology-based method in terms of both recall and precision, particularly in terms of high precision.

Compared to the morphology-based method, ResNet-18's precision, recall rate, and best F1 score improved by 0.517, 0.082, and 0.352, respectively. After implementation of the MSNN module, these values increased by 0.521, 0.199, and 0.391, respectively. Table 3 shows that compared with the ResNet-18 model, the AP improved by 0.072 after implementing the MSNN module. Furthermore, Fig. 7b shows that, after implementing the MSNN module, the optimal confidence threshold is lower, which typically results in a higher recall rate. Therefore, the results indicate that the main advantage of the MSNN module is an improvement in the recall rate.

Low gross count: the peaks in the low-count spectrum are influenced by statistical fluctuations, which in turn are affected by factors such as the radiation source intensity, detector efficiency, and measurement time. This results in the misidentification of numerous false peaks. The stronger detection capability of peaks in the low-count spectrum indicates that the proposed algorithm has enhanced anti-noise ability and can analyze low-activity radionuclide spectra more efficiently.

The fixed gross counts were set to 1×10^3 , 2×10^3 , 4×10^3 , 6×10^3 , 8×10^3 , and 2×10^4 . To reduce the impact of a low SNR, the random SNR was controlled within the range of 0.5~1, and 20,000 samples were regenerated for

Table 2 Parameters used in the morphological-based method

Steps	Smoothing		Top hat	False-peak screening
Parameters	w	k	L	W_T
Traversal range	9~49	3,5,7	4~40	2~15
Best parameters	25	5	16	14

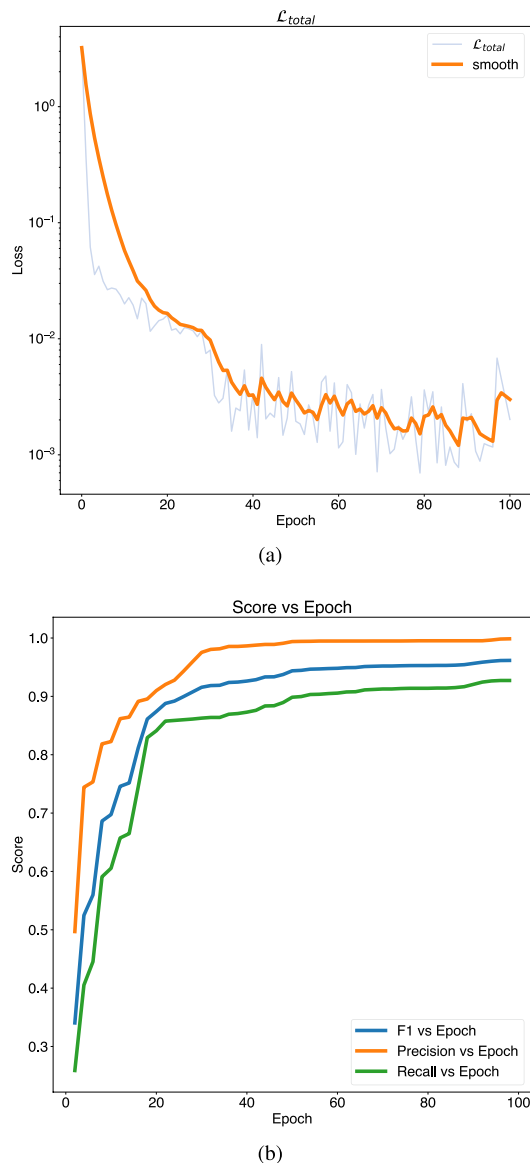


Fig. 6 (Color online) Performance of validation dataset during training; (a) loss curve during training; (b) performance of validation dataset

performance testing. The best F1 score, precision, and recall among the three methods were evaluated based on the test set.

As shown in Fig. 8, as the gross count decreased, the performance indicators of all three methods also decreased. However, the F1 score of the ResNet-18+MSNN model outperformed those of the other two methods, as shown in Fig. 8a.

As shown in Fig. 8c, the recall of the morphology-based method is less affected and matches that of the DL-based method. This is because large statistical fluctuations can

increase the false-alarm rate; however, the method can still identify existing peaks.

The increased false-alarm rate led to a rapid decrease in precision, as shown in Fig. 8b. As shown in Fig. 8c, the recall rate of the method without the MSNN module gradually decreases as the statistical fluctuations increase.

Overall, as shown in Fig. 8b, the DL-based method consistently maintained the advantage of high precision. The MSNN module effectively improved the recall rate.

Low SNR: the SNR of the spectrum indicates the prominence of the peak relative to the background, which is typically influenced by the intensity of the radioactive source's background. The stronger peak detection capability in the low SNR spectra suggests that the proposed algorithm can effectively identify low-intensity peaks buried in a strong background.

To evaluate the performance under low SNR conditions, the SNR values were set to 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, and 0.5. To minimize the impact of low gross counts, the random gross count range was controlled between $4 \times 10^4 \sim 1 \times 10^5$, and 20,000 samples were regenerated for performance testing. The impact of SNR on the three methods was evaluated (Fig. 9).

The DL-based method consistently maintained higher precision at all the SNR values. Although the F1 score decreased at low SNRs, it remained higher than that of the morphology-based method. As shown in Fig. 9c, the recall rate is significantly improved by the MSNN module, particularly for a low SNR ($SNR < 0.3$).

Radionuclide category: different radionuclide categories can represent peaks with different half-widths to a certain extent; a smaller FWHM corresponds to a smaller peak area. Small-object detection remains challenging in computer vision [43]. Previous experimental results show that when the MSNN module is applied, the improvement in the recall rate exceeds that in the precision metric. To further investigate the underlying mechanisms, we separately evaluated the detection performance of DL-based methods for peaks of varying widths.

The test set comprised samples containing only one type of radionuclide. To reduce the impact of a low gross counts and SNRs, the random gross count range was set to $5 \times 10^4 \sim 1 \times 10^5$; the SNR range was set to 0.5~1.0, and 20,000 samples were regenerated for the performance test.

Figure 10 shows the performance of the DL-based method for different radionuclides. These results are obtained by evaluating the performance of the model for narrower and broader peaks. As shown in Fig. 10c, without the MSNN module, the recall rates for the ^{241}Am and ^{57}Co peaks are significantly lower than those for the other nuclides. This result suggests that the missed detection rate is higher because of

Fig. 7 (Color online) Comprehensive performance comparison chart; **a** comprehensive performance of P-R curve; **b** comprehensive performance of F1-confidence

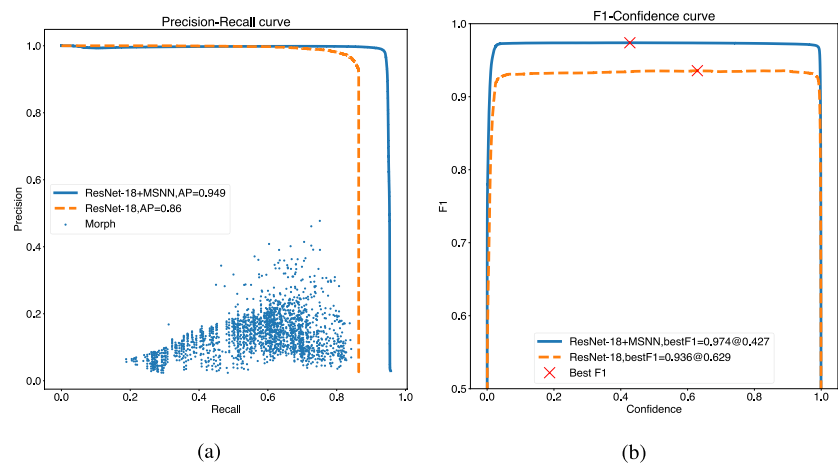


Table 3 Comprehensive performance comparison table

Method	Performance metrics			
	Best F1@confidence	Precision	Recall	AP
MorPh	0.583	0.477	0.751	–
ResNet-18	0.935@0.629	0.994	0.883	0.86
ResNet-18+MSNN	0.974@0.427	0.998	0.95	0.946

Bold formatting indicates the best (highest) value for each performance metric across the compared methods.

the small area under the narrow peaks. After applying the MSNN module, the recall rate improves significantly, which demonstrates the effectiveness of the proposed method.

As shown in Fig. 10c, the recall rate for ^{137}Cs is significantly lower than those for ^{134}Cs and ^{60}Co , despite the similar FWHM values of their peaks. This observation can be attributed to the small number of characteristic peaks of ^{137}Cs

Cs. In other words, the effective area occupied by the characteristic peak of ^{137}Cs in the complete energy spectrum is small, and the numbers of positive and negative samples are highly unbalanced.

Qualitative analysis: Figures 11 and 12 illustrate the peak search and segmentation results of the morphological methods, ResNet-18 and ResNet-18+MSNN. The morphological peak search results are indicated by red arrows. The DL model output includes the confidence (blue bar graph), characteristic peak area (red rectangle), and ground truth (gt) value, as indicated by a green dotted rectangle.

As shown in Fig. 11, the energy spectrum contains peaks derived from mixed nuclides ^{57}Co , ^{60}Co , ^{134}Cs , and ^{137}Cs , with significant overall statistical fluctuations. In addition, overlapping characteristic peaks appear at 605 keV (^{134}Cs) and 662 keV (^{137}Cs).

In this case, the morphological method generates numerous false positives, leading to reduced precision.

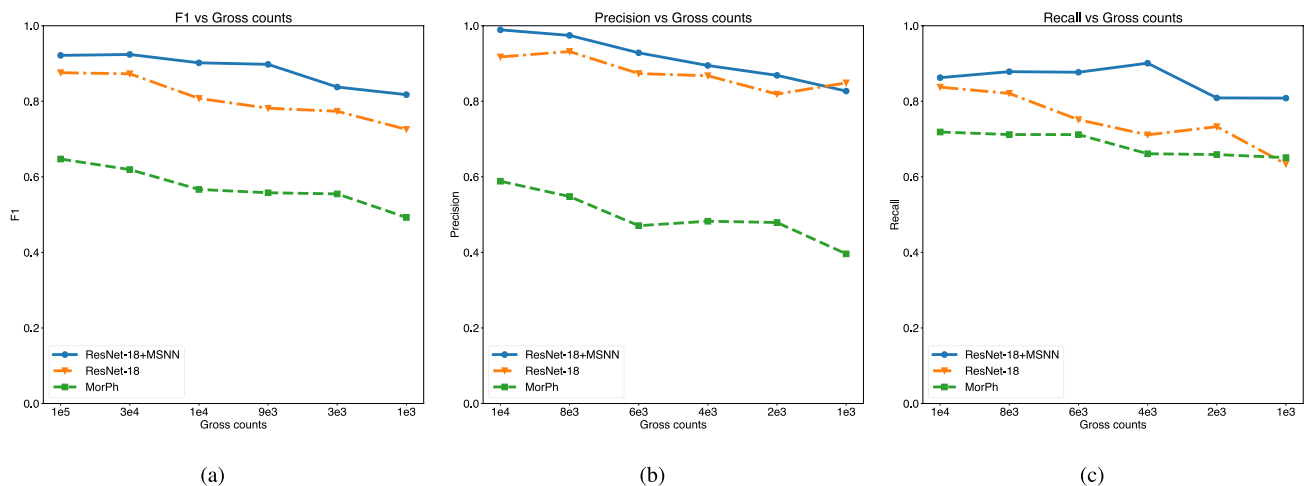


Fig. 8 (Color online) Performance of different low gross counts; **a** F1 score at different gross counts; **b** precision at different gross counts; **c** recall at different gross counts

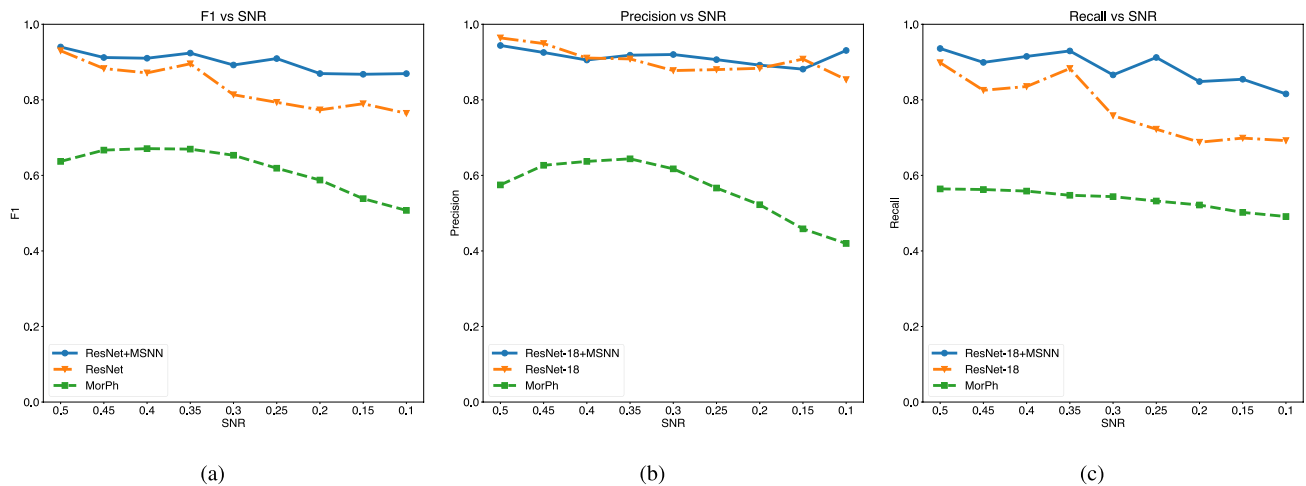


Fig. 9 (Color online) Performance of different low SNRs; **a** F1 score at different SNRs; **b** precision at different SNRs; **c** recall at different SNRs

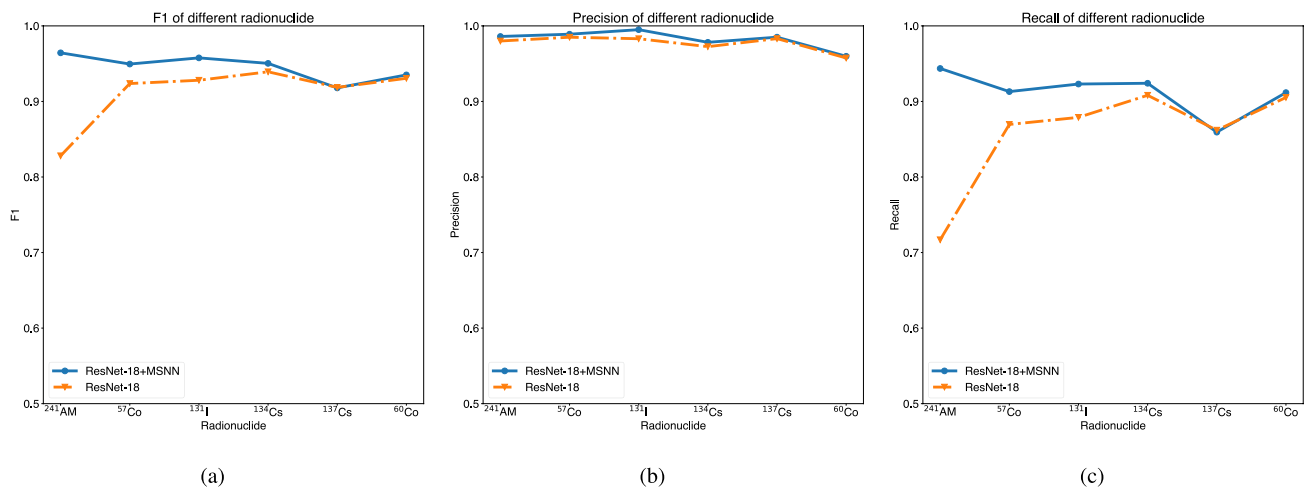


Fig. 10 (Color online) Performance of different radionuclides; **a** F1 score at different radionuclides; **b** precision at different radionuclides; **c** recall at different radionuclides

However, the ResNet-18 model does not provide sufficient confidence for the 605 keV (^{134}Cs) characteristic peak, resulting in a lower recall rate. The segmentation precision of the ResNet-18 model for the 1173 keV (^{60}Co) characteristic peak area is significantly lower than that of the ResNet-18+MSNN model.

The energy spectrum in Fig. 12 displays peaks corresponding to mixed nuclides ^{241}Am , ^{57}Co , ^{60}Co , ^{134}Cs , and ^{137}Cs . The statistical fluctuation in this energy spectrum is relatively small, with challenges imposed by the narrow peaks (59 keV for ^{241}Am) and overlapping peaks (605 keV for ^{134}Cs and 662 keV for ^{137}Cs).

The morphology-based method has a lower false-alarm rate. However, additional false-peak screening or manual parameter adjustment is required.

The DL-based method retains its high-precision advantages. However, the ResNet-18 model overlooks the characteristic peak of ^{241}Am , and the confidence prediction of the 605 keV characteristic peak of ^{134}Cs is worse than that of ResNet-18+MSNN.

The morphological method inevitably generates numerous false peaks owing to the interference of strong statistical fluctuations in the low-count full-energy peaks. The DL-based method, with its strong nonlinear fitting ability, effectively processes noise interference and maintains a high precision. The MSNN module effectively models the intricate features of the overlapping peak areas, thereby further improving the recall rate.

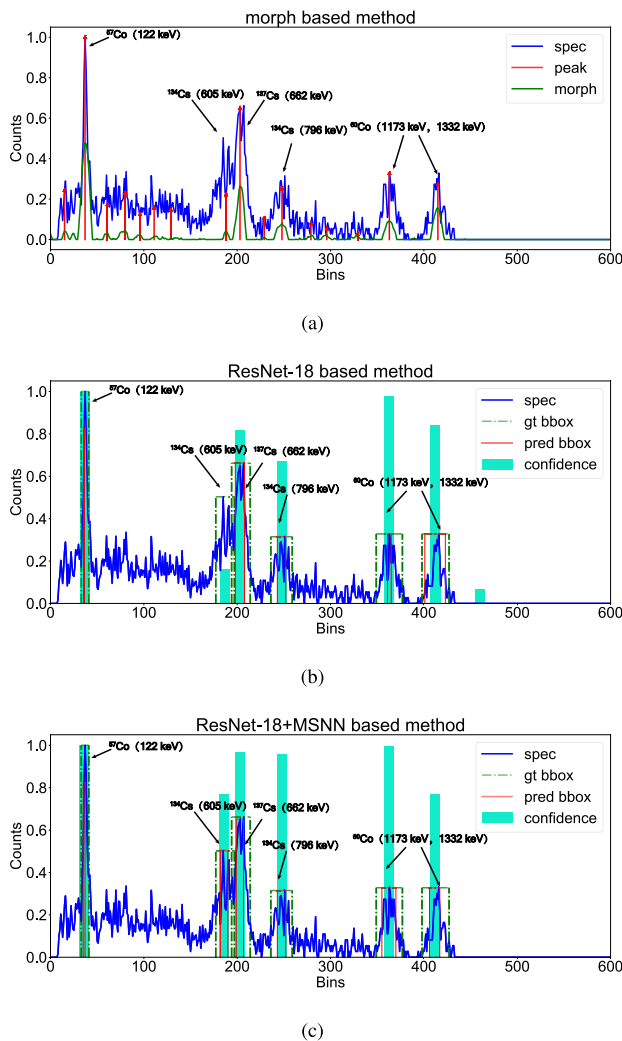


Fig. 11 (Color online) Qualitative analysis of ^{57}Co , ^{60}Co , ^{134}Cs , and ^{137}Cs mixed radionuclides; **a** MorPh-based method; **b** ResNet-18-based method; **c** ResNet-18 + MSNN-based method

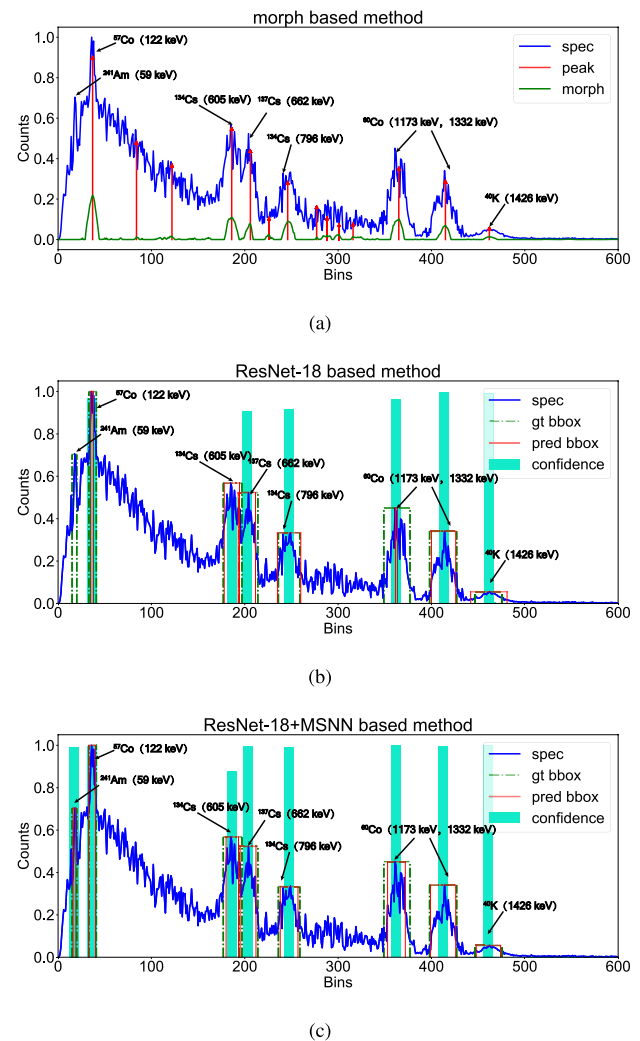


Fig. 12 (Color online) Qualitative analysis of ^{241}Am , ^{57}Co , ^{60}Co , ^{134}Cs , and ^{137}Cs mixed radionuclides; **a** MorPh-based method; **b** ResNet-18-based method; **c** ResNet-18 + MSNN-based method

Inference time: Real-time performance is a critical indicator of the applicability of practical monitoring systems. The NVIDIA Jetson AGX Orin 32GB module was used to evaluate the inference time performance, allowing users to set power consumption limits of 15, 30, 50, and 60 W. This setup enabled the evaluation of both inference time and power consumption under various application scenarios.

We conducted experiments on 60,000 samples to measure the inference times under various power consumption constraints using Python 3.8 and PyTorch 2.1. The results are presented in Table 4.

As shown in Table 4, under a 15-W power consumption limit, the average inference time per sample is 16.1941 ms, with a maximum of 112.4730 ms and minimum of 13.5004

ms. This result confirms the good real-time performance of the proposed model.

Table 4 Inference time performance

Power (W)	Inference time (ms)		
	Min	Average	Max
15	13.5004	16.1941	112.4730
30	9.5539	10.4092	30.1266
50	10.3586	11.4204	21.4953
60	7.2591	7.8072	21.9126

5 Conclusion

This paper presents YOLOspecNN—a novel method that applies computer vision object detection techniques to radio-nuclide spectrum peak search and segmentation. Our method automatically segments the full-energy peak regions of gamma spectra, without requiring manual parameter adjustments, while supporting both qualitative and quantitative analyses.

The proposed approach offers significant advantages over traditional morphology-based methods, particularly in mitigating challenges arising from complex mixed-nuclide energy spectra, low gross counts, low SNRs, and overlapping peaks. Additionally, we introduce an MSNN module that enhances the local context features and improves the recall performance for low-intensity or narrow peaks in low-energy regions. The YOLOspecNN model provides real-time inferences on a 15-W low-power embedded platform.

Full-energy peaks exhibit relatively fixed Poisson or Gaussian statistical properties, which increase the flexibility of the model and may enable better handling of various detectors and spectra. In the future, we plan to expand the dataset to include additional detector types and nuclide categories to further enhance the robustness and generalization of the model.

Author Contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Cao-Lin Zhang, Jiang-Mei Zhang, Hao-Lin Liu, Shu-Ya Qin, and Jia-Qi Wang. The first draft of the manuscript was written by Cao-Lin Zhang, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data Availability The data that support the findings of this study are openly available in Science Data Bank at <https://cstr.cn/31253.11.sciencedb.35952> and <https://www.doi.org/10.57760/sciencedb.35952>.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. J. Eberth, J. Simpson, From Ge(Li) detectors to gamma-ray tracking arrays-50 years of gamma spectroscopy with germanium detectors. *Prog. Part. Nucl. Phys.* **60**, 283–337 (2008). <https://doi.org/10.1016/j.pnpnp.2007.09.001>
2. 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)
3. 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>
4. Y. Eisen, A. Shor, I. Mardor, CdTe and CdZnTe X-ray and gamma-ray detectors for imaging systems. *IEEE Trans. Nucl. Sci.* **51**, 1191–1198 (2004). <https://doi.org/10.1109/TNS.2004.829437>
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. M. Alamaniotis, S. Lee, T. Jevremovic, Intelligent analysis of low-count scintillation spectra using support vector regression and fuzzy logic. *Nucl. Tech.* **191**, 41–57 (2015). <https://doi.org/10.13182/NT14-75>
7. M. Monterial, K. Nelson, S. Labov et al., Benchmarking algorithm for radio nuclide identification (BARNI) literature review (2019). <https://www.osti.gov/biblio/1544518>
8. 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>
9. M. Rawool-Sullivan, J. Bounds, S. Brumby, et al., *Steps Toward Automated Gamma Ray Spectroscopy* (Los Alamos National Laboratory, 2010)
10. K. Lam, W. Zhang, Gamma peak search and peak fitting algorithm for a low-resolution detector with applications in gamma spectroscopy. *J. Radioanal. Nucl. Chem.* **322**, 255–261 (2019). <https://doi.org/10.1007/s10967-019-06760-x>
11. 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)
12. 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)
13. G. Xiao, L. Deng, B. Zhang et al., A nonlinear wavelet method for data smoothing of low-level gamma-ray spectra. *J. Nucl. Sci. Technol.* **41**, 73–76 (2004). <https://doi.org/10.3327/jnst.41.73>
14. C.J. Sullivan, S.E. Garner, K.B. Butterfield, Wavelet analysis of gamma-ray spectra, in *IEEE Symposium Conference Record Nuclear Science 2004* (2004), pp. 281–286. <https://doi.org/10.1109/NSSMIC.2004.1462198>
15. C.J. Sullivan, M.E. Martinez, S.E. Garner, Wavelet analysis of sodium iodide spectra. *IEEE Nucl. Sci. Sympos. Confer. Record* **2005**(1), 302–306 (2005). <https://doi.org/10.1109/NSSMIC.2005.1596258>
16. H. Yang, X. Zhang, W. Gu et al., A novel method for gamma spectrum analysis of low-level and intermediate-level radioactive waste. *Nucl. Sci. Tech.* **34**, 87 (2023). <https://doi.org/10.1007/s41365-023-01236-w>
17. R. Luo, Y. Huang, X. Liu, Comparative study on peak-seeking methods of the mixed radioactive energy spectrum. *Open Access Library J.* **7**, 1–10 (2020). <https://doi.org/10.4236/oalib.1106676>
18. Y. Huang, M. Liu, X. Liu et al., Nuclide spectrum peak searching algorithm based on multiple morphological structuring elements. *J. Phys. Confer. Ser.* **1634**, 012078 (2020). <https://doi.org/10.1088/1742-6596/1634/1/012078>
19. H. Liu, H. Ji, J. Zhang 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>
20. H. Liu, H. Ji, J. Zhang 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>
21. J. Wang, W. Gu, H. Yang et al., Analytical method for γ energy spectrum of radioactive waste drum based on deep neural network.

- Nucl. Tech. **45**, 040501 (2022). <https://doi.org/10.11889/j.0253-3219.2022.hjs.45.040501>. (in Chinese)
22. 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**, 110212 (2022). <https://doi.org/10.1016/j.apradiso.2022.110212>
23. S. Galib, P. Bhowmik, A. 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>
24. S. Wu, X. Tang, P. Gong et al., Peak-searching method for low count rate γ spectrum under short-time measurement based on a generative adversarial network. Nucl. Instrum. Methods Phys. Res. A. **1002**, 165262 (2021). <https://doi.org/10.1016/j.nima.2021.165262>
25. R. Zhao, N. Liu, Low-resolution gamma-ray spectrum analysis using comprehensive training set and deep ResNet architecture. Nucl. Instrum. Methods Phys. Res. A. **1050**, 168135 (2023). <https://doi.org/10.1016/j.nima.2023.168135>
26. A. Voulodimos, N. Doulamis, A. Doulamis et al., Deep learning for computer vision: a brief review. Intell. Neurosci. **2018**, 7068349 (2018). <https://doi.org/10.1155/2018/7068349>
27. Z. Zou, K. Chen, Z. Shi et al., Object detection in 20 years: a survey. Proc. IEEE **111**, 257–276 (2023). <https://doi.org/10.1109/JPROC.2023.3238524>
28. A. Hafiz, G. Bhat, A survey on instance segmentation: state of the art. Int. J. Multimed. Inf. Retr. **9**, 171–189 (2020). <https://doi.org/10.1007/s13735-020-00195-x>
29. J. Redmon, S. Divvala, R. Girshick et al., You only look once: unified, real-time object detection, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), pp. 779–788. <https://doi.org/10.1109/CVPR.2016.91>
30. K. He, X. Zhang, S. Ren et al., Deep residual learning for image recognition, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), pp. 770–778. <https://doi.org/10.1109/CVPR.2016.90>
31. K. He, X. Zhang, S. Ren et al., Identity mappings in deep residual networks, in *Computer Vision—ECCV 2016* (2016), pp. 630–645
32. G. Qi, Y. Zhang, K. Wang et al., Small object detection method based on adaptive spatial parallel convolution and fast multi-scale fusion. Remote. Sens.-Basel **14**, 420 (2022). <https://doi.org/10.3390/rs14020420>
33. Z. Xin, Y. Cai, T. Lu et al., Localizing discriminative visual landmarks for place recognition, in *2019 International Conference on Robotics and Automation (ICRA)* (2019), pp. 5979–5985. <https://doi.org/10.1109/ICRA.2019.8794383>
34. H. Kim, E. Dunn, J. Frahm, Learned contextual feature reweighting for image geo-localization, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 3251–3260. <https://doi.org/10.1109/CVPR.2017.346>
35. 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)
36. J. Li, S. Liu, Y. Zhang et al., Pre-assessment of dose rates of ^{134}Cs , ^{137}Cs , and ^{60}Co for marine biota from discharge of Haiyang Nuclear Power Plant, China. J. Environ. Radioactiv. **147**, 8–13 (2015). <https://doi.org/10.1016/j.jenvrad.2015.05.001>
37. 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**, 96–104 (2013). <https://doi.org/10.1016/j.jenvrad.2012.11.009>
38. 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**, 103–113 (2012). <https://doi.org/10.1016/j.jenvrad.2012.02.006>
39. R. Padilla, S. Netto, E. Silva, A survey on performance metrics for object-detection algorithms, in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)* (2020), pp. 237–242. <https://doi.org/10.1109/IWSSIP48289.2020.9145130>
40. R. Padilla, W. Passos, T. Dias et al., A comparative analysis of object detection metrics with a companion open-source toolkit. Electronics-Switz **10**, 279 (2021). <https://doi.org/10.3390/electronics10030279>
41. H. Yule, Mathematical smoothing of gamma ray spectra. Nucl. Instrum. Methods **54**, 61–65 (1967). [https://doi.org/10.1016/S0029-554X\(67\)80007-7](https://doi.org/10.1016/S0029-554X(67)80007-7)
42. S. Islami Rad, Optimization of energy window for gamma densitometer based backscatter method in oil industry. Russ. J. Nondestruct. **52**, 245–249 (2016). <https://doi.org/10.1134/S1061830916040045>
43. Y. Liu, P. Sun, N. Wergeles et al., A survey and performance evaluation of deep learning methods for small object detection. Expert Syst. Appl. **172**, 114602 (2021). <https://doi.org/10.1016/j.eswa.2021.114602>

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.