



Reinforcement learning for muon scattering tomography enhancement

Yi-Ni Wu¹ · Yuan-Yuan Liu^{2,3,4,5} · Li Wang^{2,3,4} · Jian-Jie Zhang¹ · Ning Su^{2,3,4,5} · Wen-Wan Ding^{2,3,4,5} · Xin Zhao^{2,3,4,5} · Zhi Zhou^{2,3,4,5} · Peng Zheng^{2,3,4,5} · Jian-Ping Cheng^{2,3,4,5}

Received: 6 March 2025 / Revised: 7 July 2025 / Accepted: 11 July 2025 / Published online: 11 February 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

Muon scattering tomography (MST) is a powerful noninvasive imaging technique with significant applications in nuclear material detection and security screening. Traditional MST usually relies on the point of closest approach (PoCA) algorithm to reconstruct images from muon scattering data; however, PoCA often suffers from suboptimal image clarity and resolution. To overcome these challenges, we propose a novel approach that leverages reinforcement learning (RL) to enhance MST reconstruction, termed the μ RL-enhanced method. By framing the MST optimization task as an RL problem, we developed an intelligent agent capable of dynamically adjusting the key PoCA parameters. The agent is trained using a multi-objective reward function that guides the optimization toward higher-quality reconstructions. Our experimental results show that the μ RL-enhanced method significantly outperforms the traditional PoCA baseline across multiple benchmark metrics. Specifically, the proposed approach on average attains a 307% improvement in the intersection over union (IoU), a 79% increase in the structural similarity index measure (SSIM), and a 8.4% enhancement in the peak signal-to-noise ratio (PSNR) across four experiments. Furthermore, when benchmarked against the maximum likelihood scattering and displacement (MLSD) algorithm, the μ RL-enhanced method offers modest gains in PSNR and IoU, together with a one-third increase in SSIM. These improvements demonstrate the enhanced reconstruction accuracy and structural fidelity of the μ RL-enhanced method, highlighting its potential to advance MST technologies and their applications.

Keywords Muon scattering tomography · Reinforcement learning · Q-learning · PoCA

This work was supported by the National Natural Science Foundation of China (No.12222502).

✉ Yuan-Yuan Liu
yyluu@bnu.edu.cn

✉ Li Wang
wangl@bnu.edu.cn

¹ College of Nuclear Technology and Automation Engineering, Chengdu University of Technology, Chengdu 610059, China

² School of Physics and Astronomy, Beijing Normal University, Beijing 100875, China

³ Joint Laboratory of Jinping Ultra-low Radiation Background Measurement of Ministry of Ecology and Environment Beijing Normal University, Beijing 100875, China

⁴ Jinping Deep Underground Frontier Science and Dark Matter Key Laboratory of Sichuan Province, Xichang 615000, China

⁵ Key Laboratory of Beam Technology of Ministry of Education, Beijing 100875, China

1 Introduction

Muon scattering tomography (MST) is a noninvasive imaging technique that uses cosmic-ray muons to reconstruct the internal structure of dense materials. It is a versatile and powerful technique, particularly beneficial for probing nuclear materials [1–4]. It aids in the identification of concealed radioactive substances and illicit nuclear materials [5–7] in various environments. For example, MST plays a crucial role in nuclear waste management and storage facility inspections [8, 9], offering a reliable method for assessing the integrity of storage facilities and monitoring the containment of hazardous materials. In addition, MST is indispensable in cargo scanning and security screening, where it helps detect hidden contraband in shipments [10, 11], thus ensuring security at the ports and borders. Finally, MST is also applied in radiation shielding studies, where it assists in evaluating and improving radiation shielding

materials and techniques for both industrial and research purposes [12]. The precision and effectiveness of MST are crucial, as they directly impact the reliability of detection and analysis in these applications.

The point of closest approach (PoCA) [13] algorithm is commonly used in the MST to reconstruct muon scattering locations. Although PoCA is simple and computationally efficient, it has inherent limitations in terms of image clarity and resolution, primarily because of data sparsity and high noise levels in muon measurements.

To overcome these challenges, we propose integrating reinforcement learning (RL) [14]—a technique effective in optimizing complex, nonlinear problems through iterative interactions with the environment—into MST. By adaptively tuning the algorithmic parameters, RL can mitigate the limitations of PoCA and significantly improve the image quality, enabling precise imaging that detects subtle changes in the monitored objects.

This study introduces a new paradigm for image reconstruction and has the potential to significantly advance the field of muon scattering tomography. By demonstrating the feasibility of RL in this context, we aim to pave the way for further exploration of adaptive learning techniques in tomography and related areas. In summary, the key contributions of this study are as follows:

1. Novel integration of RL with PoCA for MST, introducing an intelligent approach to enhance muon imaging.
2. Development of a new RL framework for MST, including a multidimensional state representation, a multi-domain action space, and a reference-based multi-objective reward system.
3. Comprehensive evaluation of the μ RL-enhanced approach, demonstrating significant improvements in image quality metrics compared to traditional PoCA methods.

The remainder of this paper is organized as follows: Sect. 2 provides a background on muon scattering tomography and reinforcement learning, Sect. 3 describes the proposed methodology, Sect. 4 presents the experimental results, and Sect. 5 concludes the paper and outlines directions for future research.

2 Background

2.1 Muon scattering tomography techniques

Traditional MST algorithms aim to reconstruct muon scattering locations by analyzing deviations in muon trajectories as they pass through matter. PoCA is a foundational method used for this purpose because of its simplicity and

computational efficiency. PoCA estimates the scattering point by determining the midpoint of the line connecting the closest points between the incoming and outgoing muon tracks. Mathematically, given an incoming muon track defined by the direction vector $\mathbf{u}$ and point P_{in} , and an outgoing track with direction vector $\mathbf{v}$ and point P_{out} , the PoCA point P_{PoCA} is calculated by solving the following:

$$\min_{s,t} \left\| P_{\text{in}} + s\mathbf{u} - (P_{\text{out}} + t\mathbf{v}) \right\|, \quad (1)$$

where s and t are the scalar parameters along the tracks. Solving this minimization problem yields

$$s^* = \frac{(\mathbf{v} \cdot (P_{\text{in}} - P_{\text{out}}))(\mathbf{u} \cdot \mathbf{v}) - (\mathbf{u} \cdot (P_{\text{in}} - P_{\text{out}}))(\mathbf{v} \cdot \mathbf{v})}{(\mathbf{u} \cdot \mathbf{u})(\mathbf{v} \cdot \mathbf{v}) - (\mathbf{u} \cdot \mathbf{v})^2}, \quad (2)$$

$$t^* = \frac{(\mathbf{v} \cdot (P_{\text{in}} - P_{\text{out}}))(\mathbf{u} \cdot \mathbf{u}) - (\mathbf{u} \cdot (P_{\text{in}} - P_{\text{out}}))(\mathbf{u} \cdot \mathbf{v})}{(\mathbf{u} \cdot \mathbf{u})(\mathbf{v} \cdot \mathbf{v}) - (\mathbf{u} \cdot \mathbf{v})^2}. \quad (3)$$

The PoCA point is then given by

$$P_{\text{PoCA}} = \frac{1}{2} \left((P_{\text{in}} + s^*\mathbf{u}) + (P_{\text{out}} + t^*\mathbf{v}) \right). \quad (4)$$

The distance between the two closest points is derived as

$$d_{\text{closest}} = \left\| P_{\text{in}} + s^*\mathbf{u} - (P_{\text{out}} + t^*\mathbf{v}) \right\|. \quad (5)$$

Additionally, the angle between the incoming and outgoing muon tracks is

$$\theta = \cos^{-1}(\mathbf{u} \cdot \mathbf{v}). \quad (6)$$

Assume that the detection volume is divided into M voxels and that there are N_j scattering points in the j_{th} voxel. The density map D is calculated as

$$D_j = \sqrt{\frac{\sum_{i=1}^{N_j} \theta_i^2}{N_j}}, \quad j = 1, \dots, M. \quad (7)$$

In contrast to the basic PoCA strategy, the second strategy incorporates a weighted approach based on the angles between the incoming and outgoing muons. This approach addresses the scenarios in which the scattering point may not be exactly halfway between the two tracks, particularly when the muon trajectories diverge significantly. The angle-based weight w_i is defined as

$$w_i = \frac{\theta_i}{\theta_{\text{max}}}, \quad (8)$$

where θ_{max} is the maximum scattering angle. The scattering point in this case is determined by

$$P_{\text{scatter}} = w_i(P_{\text{in}} + s^* \mathbf{u}) + (1 - w_i)(P_{\text{out}} + t^* \mathbf{v}). \quad (9)$$

This method assigns greater importance to the incoming muon trajectory with a larger scattering angle, as this results in the scattering point being closer to the incoming direction of the muon. The angle-weighted strategy tends to make the reconstruction process more reliable in situations where the incoming and outgoing muon tracks have large scattering angles, as verified in our work.

Despite its speed, PoCA's single-scatter assumption smears edges and introduces artifacts once a muon undergoes multiple Coulomb interactions inside complex high-Z materials. Schultz et al. (2003) therefore introduced two statistical methods: (i) maximum likelihood scattering (MLS), which uses only the scattering angle, and (ii) maximum likelihood scattering and displacement (MLSD), which utilizes both the scattering angle and lateral shift for accurate reconstruction, although the scale of voxel-wise updates increases with resolution [13, 15]. Liu et al. (2007) subsequently developed EM implementations for the scattering only model (MLS-EM) and the joint scattering and displacement model [16]; by introducing an ordered subsets scheme, they further sped these algorithms up to produce MLS-OSEM and MLSD-OSEM [17]. To curb noise without losing detail, Wang et al. (2007) reformulated the MLSD within a maximum a posteriori (MAP) framework, adding Gaussian (smooth) and Laplacian (sparse) priors and applying voxel-wise shrinkage after each step; MAP suppresses high-intensity noisy voxels, preserves edges, and remains robust even with low flux data [18]. Further refinement using the Gaussian scale mixture (GSM) likelihood matches the heavy-tailed scattering statistics with a weighted Gaussian mix, yielding cleaner, high-Z-focused images [19]. Together, PoCA, MLS, MLSD, MAP, and GSM span a spectrum of reconstruction methods, from PoCA speed to full Bayesian accuracy. Additionally, neural networks [20, 21] and support vector machines [22] have been applied to improve the muon reconstruction and material discrimination accuracy.

Compared to the aforementioned methods, which often require extensive iterations and detailed parameter tuning, RL can rapidly converge in complex, high-dimensional spaces, which is a more flexible and efficient high-dimensional optimization and decision-making approach. Its core advantage lies in the utilizing of real-time feedback to dynamically adjust parameters or optimization strategies to achieve optimal imaging quality. This adaptability and scalability position RL as a key factor for further enhancing the performance of MST and paving the way for more intelligent and efficient imaging systems.

2.2 Basics of reinforcement learning

Reinforcement learning is a subfield of artificial intelligence, where an agent learns to make decisions by interacting with an environment to maximize a cumulative reward over time through trial-and-error exploration [14]. Unlike supervised learning, in which the model is trained on labeled data, RL involves learning from the consequences of actions performed in an environment. In RL, an agent explores its environment, performs actions, and receives feedback in the form of rewards or penalties, which it uses to improve its future decision making.

The core components of RL are as follows:

- **Agent:** The decision maker that interacts with the environment. The agent takes action based on its current state and learns from the feedback it receives.
- **Environment:** The external system in which the agent operates. The environment defines the rules of interaction and determines the consequences of the agent's actions.
- **State (s_t):** A representation of the environment at a specific point in time. It contains all the environmental information that the agent requires to make decisions regarding its actions. The state changes over time as the agent interacts with the environment.
- **Action (a_t):** A decision or move made by the agent that impacts the environment. Actions can be discrete or continuous, depending on the problem at hand.
- **Reward (r_t):** The feedback the agent receives after taking an action in a particular state. Rewards are typically scalar values, with positive rewards indicating favorable outcomes and negative rewards indicating undesirable outcomes.
- **Policy (π):** A strategy that maps states to actions. The policy defines the agent's behavior and decides which actions to take based on the current state.
- **Transition dynamics ($P(s_{t+1} | s_t, a_t)$):** Transition dynamics define the likelihood of reaching a new state s_{t+1} given the current state s_t and action a_t . This represents the probability of transitioning to state s_{t+1} when action a_t is performed in state s_t .

In the case of MST, we aimed to optimize the PoCA reconstruction process by adjusting various parameters (actions) based on the observed image quality (state) and guiding these adjustments using a reward function that measures the improvement in reconstruction accuracy.

RL has various algorithms [23], and here we mention two of them. One common approach is Q-learning, in which the agent updates its Q-values (state-action-value function) based on feedback from the environment and seeks the action that maximizes long-term rewards. Another approach is the Deep Q-Networks (DQN), which uses deep neural

networks to approximate the Q-function, enabling the agent to handle complex, high-dimensional environments. In this study, we used Q-learning because of its simplicity.

2.3 Reinforcement learning for MST enhancement

2.3.1 Motivation

Advancements in machine learning, particularly reinforcement learning, have opened new possibilities for the optimization and enhancement of physical imaging systems. These techniques are particularly effective for adapting to dynamic conditions and refining strategies based on feedback from limited data. This adaptability is particularly valuable for systems that deal with complex, high-dimensional datasets. RL has proven to be a powerful tool across a wide range of imaging applications, including medical imaging [24, 25], computer vision [26–28], and industrial inspection. In medical imaging, RL has been applied to tasks such as landmark localization [29, 30], object detection [31], image segmentation [32, 33] and registration [34, 35]. It offers solutions to complex optimization problems, such as parameter tuning, strategy selection, and automation of diagnostic processes. For example, RL agents have been trained to adjust imaging parameters in real-time to minimize radiation exposure while maintaining image quality [36]. In the realm of computer vision, RL has been utilized for object recognition [37], image enhancement and denoising [38], as well as image generation and restoration [39]. In industrial inspection, RL has enabled the development of intelligent systems that dynamically control inspection equipment, detect defects, and optimize quality control processes [40].

In the case of MST, traditional PoCA algorithms often face challenges in terms of image clarity and resolution, primarily because of data sparsity and high noise levels inherent in the muon scattering measurements. Meanwhile, PoCA and improved PoCA algorithms typically rely on fixed parameter settings [41, 42], which are often based on prior knowledge or expert judgment, and lack the flexibility to adapt to changing conditions. Moreover, fixed parameters may struggle to handle variations in the muon flux or density distribution of target objects, particularly when dealing with more complex or heterogeneous scanned objects. By integrating RL into the MST reconstruction process, we aim to bring similar advancements to the field of muon scattering tomography, such that key PoCA parameters can be dynamically adjusted, enhancing the accuracy and efficiency of the imaging process.

2.3.2 Challenges in RL for MST

While RL has demonstrated significant success in the aforementioned imaging fields, its use for MST has not been

explored to the best of our knowledge and faces unique challenges. One of the main difficulties arises from the sparse and noisy nature of muon imaging data, which often leads to incomplete or weak signal. Unlike other imaging domains, MST is further complicated by the stochastic interactions of cosmic-ray muons, making it difficult to reliably track their scattering and accurately reconstruct the internal structures of the objects. Additionally, RL algorithms typically require substantial computational resources owing to the iterative nature of learning through repeated interactions with the environment, which poses a limitation for MST, particularly in real-time scenarios. Moreover, developing an RL framework capable of managing the intricacies of the MST reconstruction is particularly challenging. For example, the reward function must effectively quantify the image quality in the MST, as existing metrics may not fully capture the unique characteristics of muon scattering. Therefore, these frameworks must not only integrate domain-specific knowledge of muon interactions but also be sufficiently efficient to handle noisy, sparse, and sometimes ambiguous data, ensuring that the agent can learn meaningful strategies for improving reconstruction accuracy.

In the next section, we describe the MST as an RL optimization problem and the implementation of our framework in detail.

3 Methodology

3.1 MST as an RL optimization problem

MST reconstruction requires the adjustment of multiple parameters, and optimizing these parameters is challenging, particularly with varying material compositions and measurement conditions. To address this, we modeled the MST reconstruction as an RL optimization problem, enabling the adaptive refinement of parameters based on continuous feedback from the image quality.

In this framework, the environment is defined by the scattering points calculation and density-mapping process according to the PoCA algorithm, which generates updated images from muon scattering data. The RL agent interacts with the environment by adjusting the parameters to improve the image quality over time. The agent learns the optimal configurations through feedback on each reconstruction, progressively enhancing the reconstruction process.

At each iteration, the agent observes the current state, which is represented by a set of image characteristics, along with feedback defined on a set of image quality metrics. The agent then selects an action according to an ϵ -greedy exploration and exploitation policy as follows: After executing the action, the agent receives a reward and updates its Q-values. As these Q-values change, the policy is refined to

maximize long-term improvements. Through trial and error, the agent explores different parameter configurations, learns from the feedback, and adjusts its strategy. Over time, the agent converges to a policy that consistently improves the PoCA reconstruction.

For simple state-action spaces, classic methods such as Q-learning can be used, whereas more complex scenarios may require methods such as Deep Q-Networks (DQN) or actor-critic algorithms.

In summary, framing MST as an RL optimization problem allows for dynamic and adaptive parameter adjustments, thereby improving the image quality and structural accuracy.

3.2 System architecture

In this section, we develop the RL framework by leveraging a Q-learning agent to dynamically adjust the key parameters of the PoCA algorithm and post-processing steps, thereby optimizing the reconstruction process and enhancing the image quality in MST. The primary components of this framework are the action space, state representation, reward function, and agent. Each component is elaborated below to illustrate how they collaboratively facilitate effective optimization of the model.

3.2.1 Reinforcement learning framework

The system flow diagram of the RL framework is shown in Fig. 1. The process is split into two parts, as depicted by the two boxes: the main process (left side) and the Q-learning process (right side).

- Main process:
 - The main loop starts by loading the necessary input files for an experiment and then calculates the initial density map. It sets up the PoCA environment and initializes the Q-learning agent for the subsequent steps.
 - The Q-learning process is described in detail below.
 - The optimized parameters, and resulting density map, and 2D projections are saved.
- Q-learning process:
 - The agent begins by initializing the Q-table and setting the hyperparameters such as α , γ , and ϵ .
 - At each time step, the agent extracts state features from the current density map, including μ , σ^2 , γ_{1t} , and γ_{2t} . These features are then discretized to form the current state, s_t .
 - The agent uses an ϵ -greedy policy to choose an action: either a random action (with probability ϵ)

or the action with the highest expected reward (with probability $1 - \epsilon$).

- The chosen action is executed, which involves adjusting parameters.
- The environment updates the density map, and the agent receives a reward based on factors which reflect the quality of the image.
- The Q-table is updated using the Q-learning update rule, allowing the agent to learn and refine its policy.
- As the process continues, the agent decays ϵ and progressively improves its strategy for optimizing the reconstruction.

This Q-learning process loop continues until the maximum number of episodes is reached, at which point the agent policy has been optimized and the final parameters are output.

3.2.2 State, action, and reward design

The **state** representation is crucial for the agent to make informed decision making. Within the RL framework, the state was designed to capture the current state of the density map. Instead of relying solely on raw pixel values, we extracted higher-level statistical features, such as the mean, variance, skewness, and kurtosis. These features offer a comprehensive view of image quality: The mean represents the overall intensity, variance reflects the spread of the mean intensity, and skewness and kurtosis capture asymmetry and the presence of extreme values (tailing) in the image intensity distribution, respectively. The state s_t at any given time t of the current density map D_t is represented as a feature vector:

$$s_t = (\mu_t, \sigma_t^2, \gamma_{1t}, \gamma_{2t}), \quad (10)$$

where μ_t represents the average intensity of the density map, indicating the overall brightness of the reconstructed image. σ_t^2 quantifies the spread of the pixel intensity values, reflecting the noise level and structural variations. γ_{1t} measures the asymmetry of the intensity distribution, revealing whether the map is biased toward unusually low or high densities. γ_{2t} reflects how heavy-tailed or sharply-peaked the distribution is. A high kurtosis can signal either very sharp peaks or rare extreme voxels, both of which deviate from Gaussian behavior.

The **action** space consists of a discrete set of actions that the agent can take to adjust the PoCA algorithm and subsequent post-processing steps. The actions are grouped into four categories, each targeting different aspects of the reconstruction process:

a. Adjust maximum scattering distance $d_{\max}$:

- $a_0 : d_{\max} \leftarrow d_{\max} + \Delta d,$

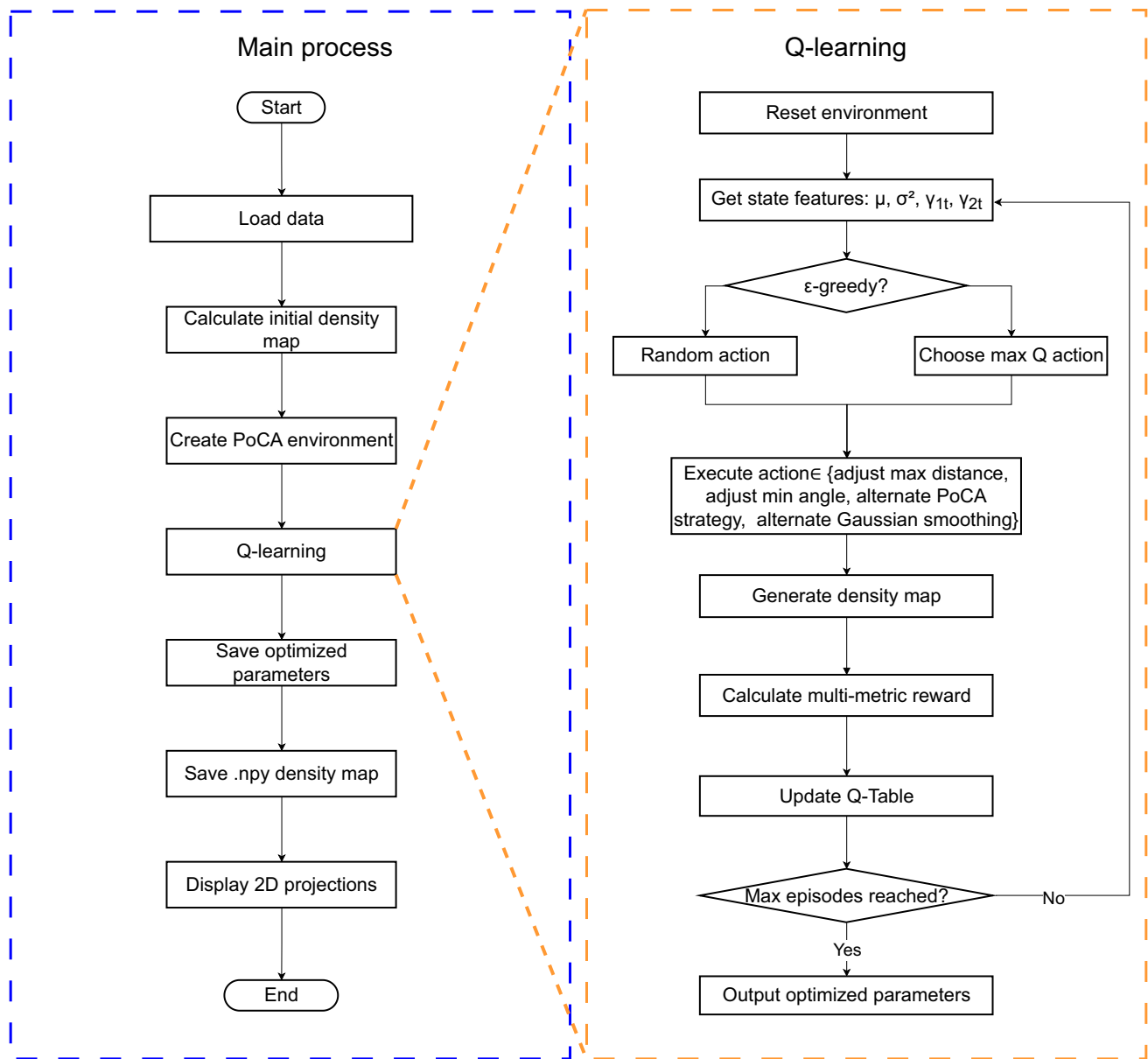


Fig. 1 (Color online) μ RL-enhanced system flow diagram

- $a_1 : d_{\max} \leftarrow d_{\max} - \Delta d.$

The agent may increase or decrease $d_{\max}$ to determine the most suitable distance range for the PoCA calculation by selecting valid scatter points with $d_{\text{closest}} \leq d_{\max}$. This parameter influences the amount of muon trajectory data used to compute the density map, thereby affecting the density map resolution.

b. Adjust minimum scattering angle $\theta_{\min}$:

- $a_2 : \theta_{\min} \leftarrow \theta_{\min} + \Delta\theta,$
- $a_3 : \theta_{\min} \leftarrow \theta_{\min} - \Delta\theta.$

The agent adjusts the minimum scattering angle threshold to filter the scattering events and improve the clarity of the reconstructed image. Increasing $\theta_{\min}$ excludes small-angle scattering (often noise), whereas decreasing it preserves subtle scattering details crucial for objects with heterogeneous internal structures.

c. Switch PoCA strategy:

- $a_4 : \text{Strategy} \leftarrow \text{Alternate PoCA strategy}.$

Different strategies exist for calculating scattering points, such as midpoint calculation versus angle-weighted

calculation. This action alternates between the two strategies to optimize the reconstruction of various materials and imaging objects.

d. Toggle Gaussian smoothing:

- a_5 : Alternate Gaussian smoothing.

This action alternates between enabling and disabling Gaussian smoothing on the density map. Smoothing can help reduce noise but may also blur the significant features. This decision depends on the need to balance noise reduction with the preservation of significant features.

Thus, the full action set is

$$\mathcal{A} = \{a_0, a_1, a_2, a_3, a_4, a_5\}.$$

By selecting appropriate actions from the action space $\mathcal{A}$, the agent can dynamically adjust the processing parameters to enhance the reconstruction quality of the density map.

The **reward** function is central to guiding the agent's learning process by evaluating improvements in image quality at each step. It considers multiple image quality metrics: clarity (edge sharpness), contrast (intensity difference), entropy (amount of uncertainty or randomness), and variance (spread of pixel intensities). By structuring the reward function in this multi-objective manner, the RL agent can simultaneously enhance various features of the reconstruction. If a particular action results in sharper features, greater contrast, and reduced noise, the agent receives a higher reward, reinforcing the effectiveness of the action and improving future decision making. We now break down each element of the reward function in detail.

a. Clarity reflects the sharpness and well-defined nature of the features in the image. A clearer image has well-defined boundaries and structures, making it easier to interpret. This can be quantified by measuring the edge strength in the density map. A suitable measure is the edge detection Sobel algorithm, formulated as

$$\begin{aligned} C(D) = & \frac{1}{N_{XY}} \sum_{i,j} \sqrt{[G_x^{XY}]_{ij}^2 + [G_y^{XY}]_{ij}^2} \\ & + \frac{1}{N_{XZ}} \sum_{i,k} \sqrt{[G_x^{XZ}]_{ik}^2 + [G_z^{XZ}]_{ik}^2} \\ & + \frac{1}{N_{YZ}} \sum_{j,k} \sqrt{[G_y^{YZ}]_{jk}^2 + [G_z^{YZ}]_{jk}^2}, \end{aligned} \quad (11)$$

where N_{XY} is the total number of pixels in the XY projection, G_x^{XY} and G_y^{XY} denote the horizontal and vertical gradients, respectively, and the indices i and j refer to the row and column pixel positions, respectively. The terms for the XZ and YZ projections are defined analogously.

b. Contrast refers to the difference in intensity between objects and the background in the image. Higher contrast can make features more distinct. The contrast $T(D)$ is as follows:

$$T(D) = \frac{D_{\max} - D_{\min}}{D_{\max} + D_{\min}}. \quad (12)$$

c. Entropy measures the amount of uncertainty or disorder in the image. High entropy indicates that the pixel values are more diverse and unpredictable, which means that the image contains a lot of detail or complexity. Low entropy indicates that the pixel values are more uniform or predictable, which means that the image is simpler or contains less detail. The Shannon entropy $H(D)$ is calculated as follows:

$$H(D) = - \sum_{i=1}^L p_i \log_2(p_i), \quad (13)$$

where $L = 256$ in our case, and p_i is the probability of the i -th histogram bin.

d. Variance measures the dispersion of pixel values. The definition we use is:

$$V(D) = \frac{1}{N} \sum_{i=1}^N (D_i - \bar{D})^2. \quad (14)$$

Ideally, the variance should be low for homogeneous images (e.g., background) and higher for images with distinct features. Because reconstruction is never perfect, we assigned a positive reward to the variance. However, a high variance may indicate artifacts in the reconstruction and should be penalized, which we do not consider in this study.

Each component contributes differently to the final image quality; therefore, they must be weighted accordingly in the reward function. By combining these factors, the agent can be guided to produce an image that is not only clear and sharp but also has a good contrast and appropriately distributed pixel intensities. To assess the improvements, the current density map was compared with the initial reference map from PoCA. The enhancement was quantified by evaluating the extent to which the new density map improved the key quality metrics. The reward can be expressed as

$$R = w_C \cdot \Delta C + w_T \cdot \Delta T + w_H \cdot \Delta H + w_V \cdot \Delta V, \quad (15)$$

where ΔC is the clarity enhancement, ΔT is the contrast improvement, ΔH is the entropy increment, ΔV is the variance growth, and $w_C = 0.4$, $w_T = 0.3$, $w_H = 0.2$, $w_V = 0.2$ are the corresponding empirical weights for each component.

3.2.3 Agent design

The RL agent in our framework employs the Q-learning algorithm [14, 43], which is a model-free RL method that aims to learn the optimal action-value function $Q(s, a)$. This function estimates the expected cumulative reward of taking action a_t in state s_t and then following the optimal policy. The Q-learning update rule is defined as:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[r_t + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t) \right], \quad (16)$$

where α is the learning rate, r_t is the reward received after taking action a_t in state s_t , γ is the discount factor, s_{t+1} is the subsequent state, and a_{t+1} represents the possible future actions.

The agent employs an ϵ -greedy policy to balance exploration and exploitation during the learning process. With probability ϵ , the agent selects a random action from the action space $\mathcal{A}$, promoting exploration of the environment. With probability $1 - \epsilon$, the agent exploits its current knowledge by choosing the action with the highest Q-value in the current state s_t , thus focusing on the actions it believes will yield the greatest reward. Over time, ϵ is gradually reduced, encouraging the agent to shift from exploration to exploitation as it becomes more confident in its learned policy. This adaptive approach ensures that the agent can discover promising new actions while simultaneously refining its decision making based on the accumulated knowledge (Fig. 5).

The agent continuously updates its policy by comparing the observed rewards associated with different actions. The reward structure is designed to reinforce actions that lead to improved image quality while penalizing actions that degrade performance. This feedback-driven approach enables the agent to learn an optimal policy $\pi(a|s)$ that maximizes the expected cumulative reward, resulting in improved MST reconstruction through adaptive parameter adjustment (Fig. 6).

4 Experiments and results

4.1 Experimental setup and MST reconstruction

The dataset utilized in this study comprises muon scattering data acquired using a muon tracking system, the schematic diagram of which is shown in Fig. 2. Eight plastic scintillator arrays are central to the tracking system and are organized into four position detectors, each featuring two orthogonal layers— x and y . This dual-layer configuration ensures precise spatial positioning, thereby enabling the

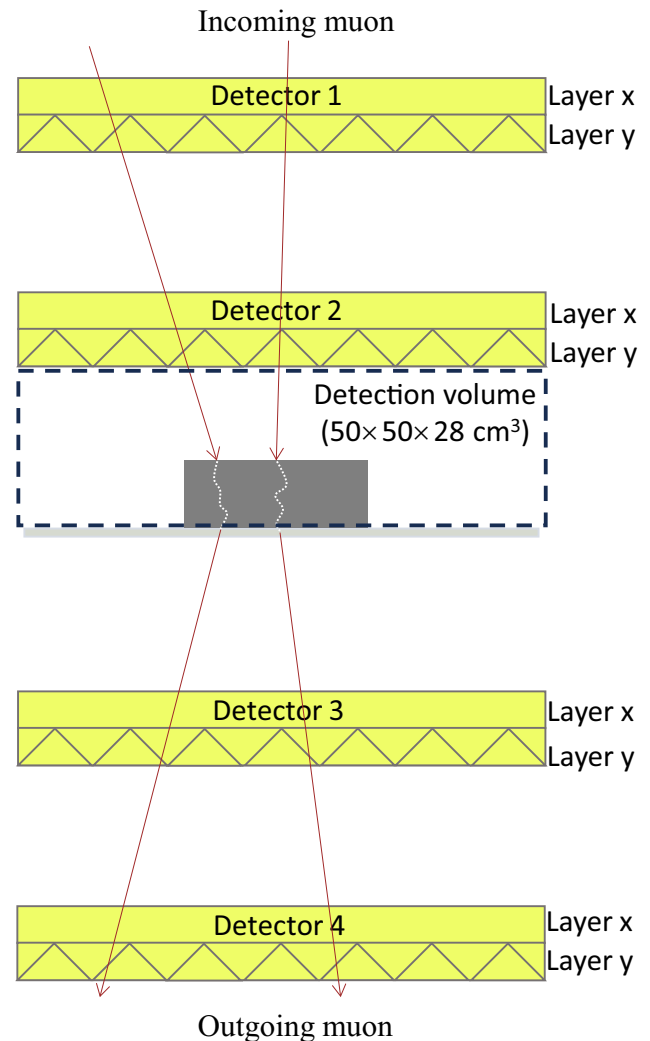


Fig. 2 (Color online) The experimental apparatus schematic diagram

accurate tracking and reconstruction of muon events. In this configuration, the tracking system encompasses a substantial detection volume of $500 \text{ mm} \times 500 \text{ mm} \times 280 \text{ mm}$, with two groups of detectors positioned above and below. This arrangement facilitates comprehensive coverage and enhances the system capability to capture muon scattering events.

Figure 3 presents the geometry layouts of the experiments, each conducted over a 24-hour period. Panel (a) shows low-, medium-, and high-Z materials using aluminum, copper, lead, and tungsten blocks. Panel (b) focuses on a hollow tungsten cylinder to study the imaging of hollow structures. Panel (c) combines a hollow tungsten cylinder with three smaller tungsten cylinders to examine the imaging of concealed and small high-Z objects. Finally, Panel (d) features 18 lead blocks arranged on three sides to form a fence, a tungsten block on the fourth side, and a central hollow

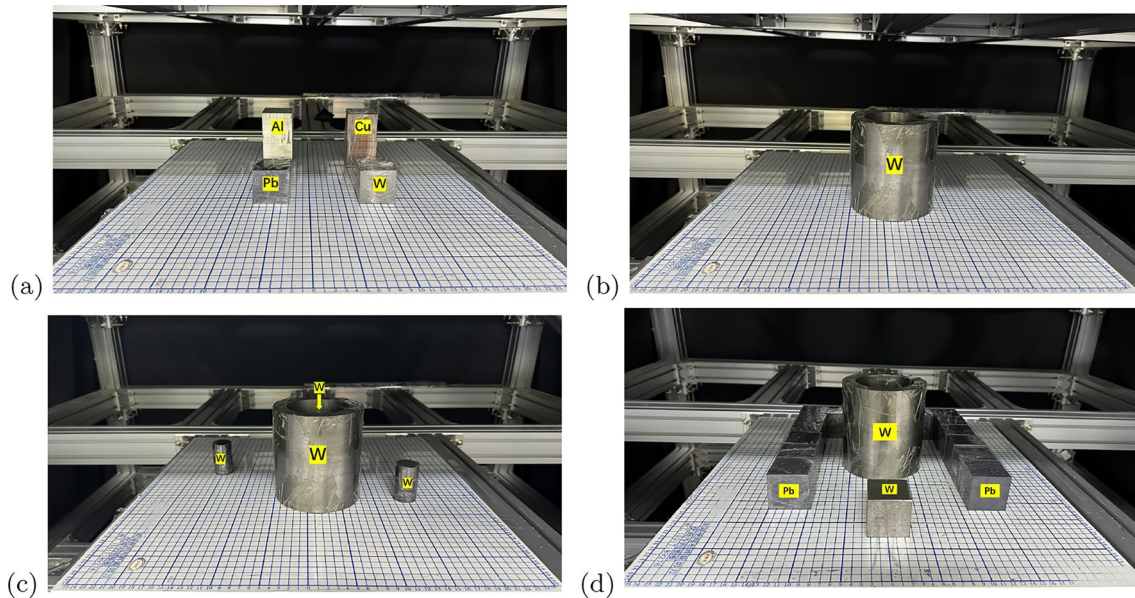


Fig. 3 (Color online) Geometry layout of the experiments. **a** Four metal blocks composed of aluminum, copper, lead, and tungsten. **b** A hollow tungsten cylinder. **c** Three small tungsten cylinders arranged with one positioned at the center of a larger hollow tungsten cylinder

tungsten cylinder, highlighting the impact of different materials and shapes on imaging performance.

Data acquisition is facilitated by synchronized scintillator signals that act as triggers for muon events. Raw data undergo preprocessing, including noise filtering, normalization, position and trajectory reconstruction, and muon track selection, ensuring high-quality input for the imaging algorithms of the traditional PoCA method and the μ RL-enhanced framework. All computing tasks were performed on a platform running Ubuntu 18.04, which was equipped with an NVIDIA GeForce RTX 2080 (8 GB GDDR6) GPU and CUDA 12.4. The computations were implemented using PyTorch 2.4.1, and data processing and visualization were supported by libraries such as NumPy, SciPy, and Matplotlib.

The RL agent employs a Q-learning variant with a learning rate α of 0.001, a discount factor γ of 0.95, and an ϵ -greedy exploration strategy starting at $\epsilon = 1.0$. The training process was set to run for up to 1000 episodes; however, empirical results showed that convergence typically occurred within 200 episodes, enabling faster training without sacrificing performance. Each episode consisted of 10 steps. Intermediate checkpoints were saved at each episode to monitor the agent's learning progress and ensure stability. Multiple trials with different parameters validated the robustness and efficiency of the RL approach, with the agent outperforming the baseline PoCA methods by adaptively selecting the optimal parameters.

and the other two placed on either side. **d** Lead blocks arranged in three sides, a tungsten block positioned on the fourth side, and a hollow tungsten cylinder situated at the center

4.2 Visual comparisons

In Figs. 4, 5, 6, and 7, side-by-side comparisons of the images produced by PoCA, MLSD, and the μ RL-enhanced method are illustrated. In the first phantom, which contains four cubes, the μ RL version clearly outlines three blocks with crisp rectangular boundaries in the XY view and shows well-separated vertical columns in the XZ and YZ planes, whereas PoCA renders only blurred clusters and MLSD blurs the cube corners. The second and third phantoms feature a hollow ring. PoCA barely hints at a ring, while MLSD recovers a faint but radially blurred ring. In contrast, the μ RL-enhanced projections exhibit a continuous ring with a clearly empty center, and the side view locates the ring axis at the correct height as the agent has learned to suppress the vertical streaking. For the final phantom, the μ RL method simultaneously preserves the thin outer frame, inner ring, side cube, and central void, whereas MLSD over smooths the frame corners, and PoCA leaves heavy salt-and-pepper noise that obscures the geometry. Along Z, the μ RL images maintain the high-density areas distinctly separated, eliminating the patchy artifacts visible in the PoCA and the diffuse haze in the MLSD. Overall, the μ RL-enhanced reconstructions present sharper edges, greater contrast between high- and low-density regions, and markedly fewer background artifacts. They produce more faithfully three-dimensional geometries than PoCA and avoid the over-smoothing that attenuates critical details in MLSD,

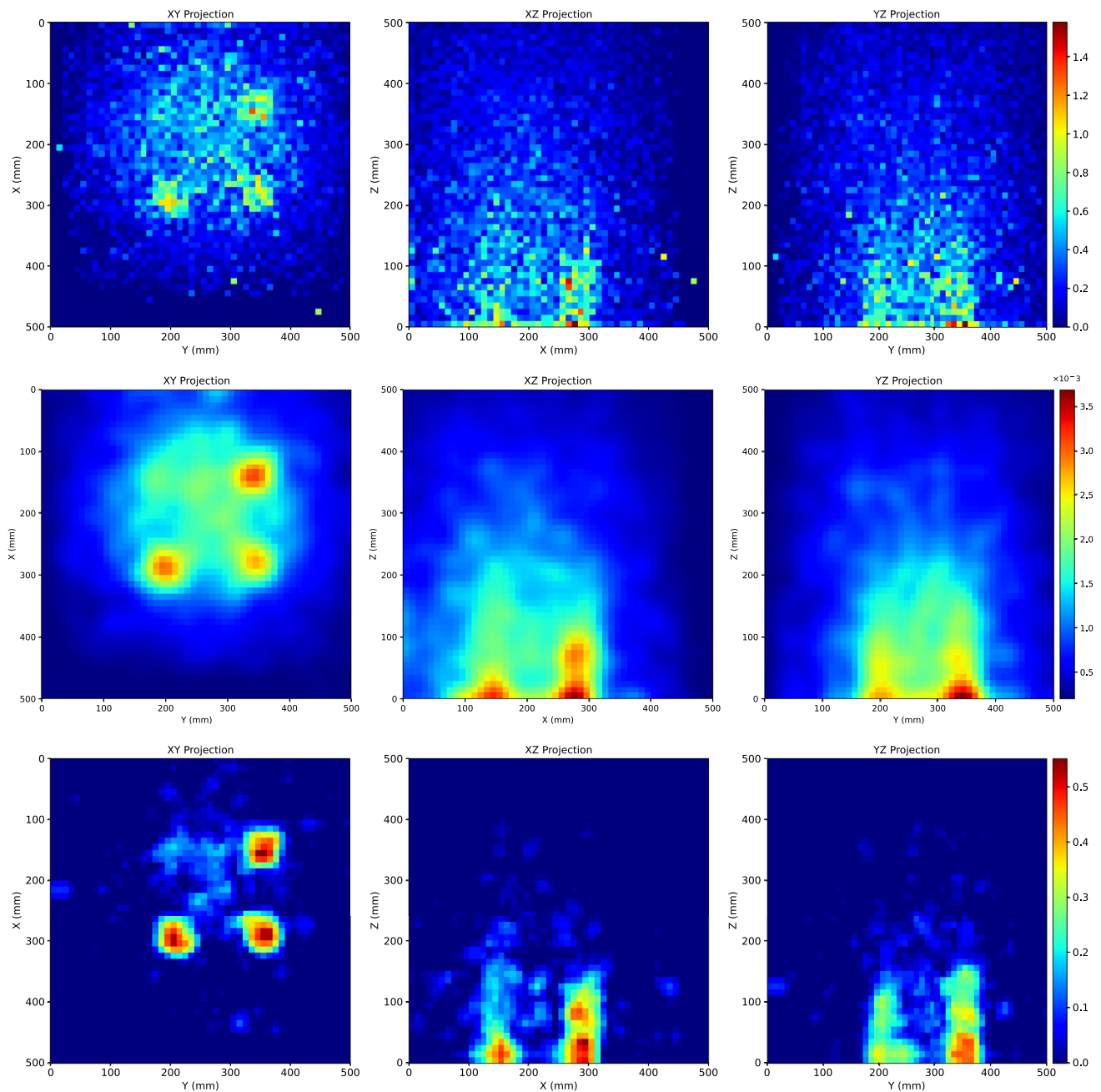


Fig. 4 (Color online) Comparison of the PoCA, MLSD, and μ RL-enhanced methods for the layout in Fig. 3a. The XY, XZ, and YZ projections obtained by the PoCA, MLSD, and μ RL-enhanced methods are shown in the upper, middle, and lower panels, respectively

providing visually and quantitatively superior images for downstream interpretation.

The four plots in Fig. 8 illustrate the evolution of the average total reward during training for the four different experimental scenarios. Each plot demonstrates a clear and steady progression of the reward, with a generally smooth and consistent increase over time. This smoothness indicates that the RL algorithm effectively learned and improved during the training process. The steady nature of

the curves reflects the stability of the training procedure, suggesting that the model gradually adapts to the environment and makes optimal adjustments to maximize the total reward. The lack of significant fluctuations or sharp changes further confirms the robustness of the training process, where the agent converges toward a reliable policy. This trend is particularly important in the context of RL, where achieving a smooth and consistent improvement in rewards typically signifies that the model is learning in

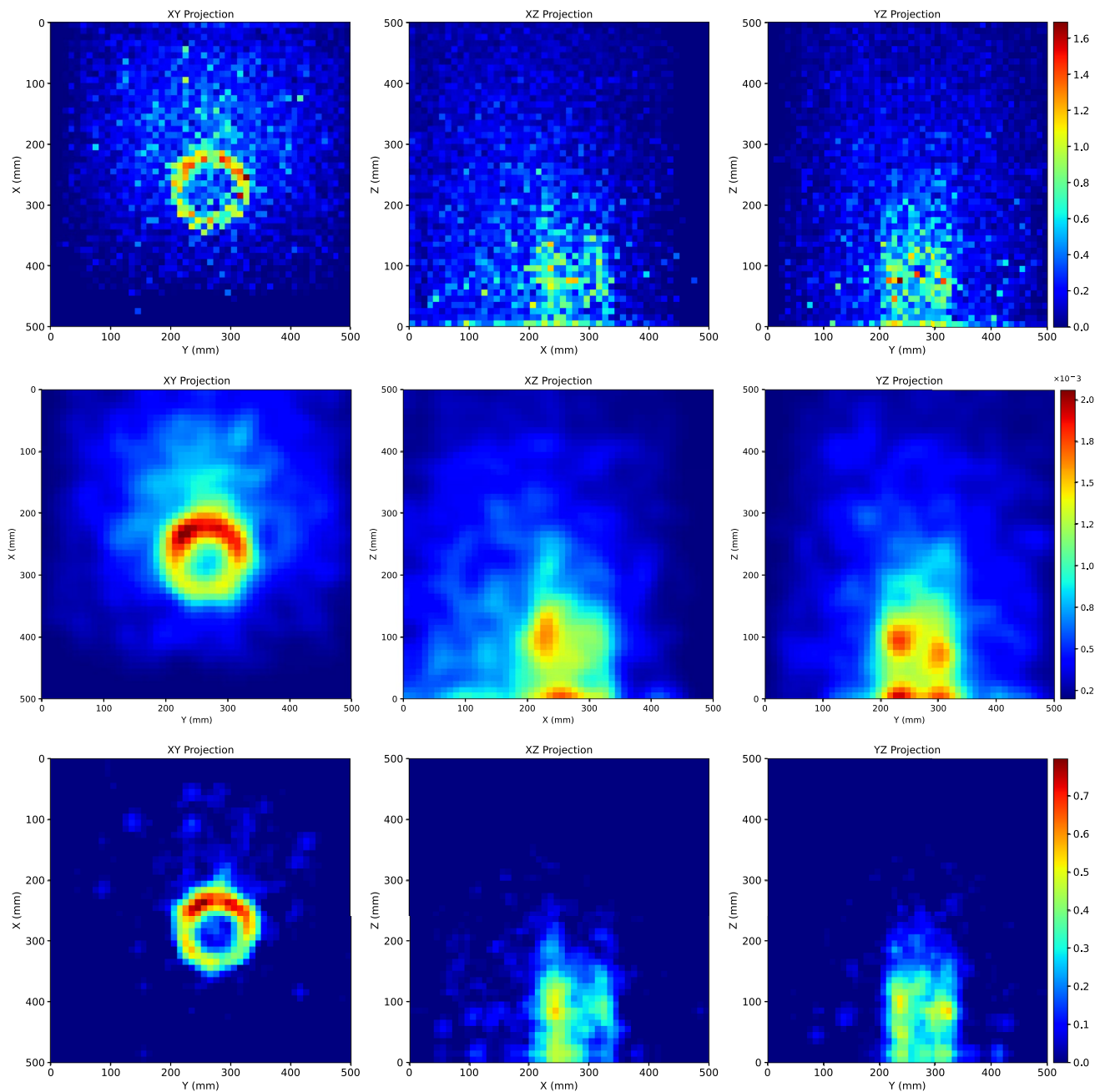


Fig. 5 (Color online) Comparison of the PoCA, MLSD method, and the μ RL-enhanced approach for the layout presented in Fig. 3b. The upper panel displays the XY, XZ, and YZ projections using PoCA, and

the middle panel shows the corresponding projections using MLSD, while the lower panel shows the corresponding projections using the μ RL-enhanced method

a meaningful way and is not overfitting or getting stuck in suboptimal solutions.

In summary, the μ RL-enhanced method provides a substantial improvement in image quality, offering sharper and more defined projections with higher contrast and better detail preservation. These advancements make it a more effective tool for capturing and analyzing complex 3D structures, addressing many of the limitations of traditional PoCA-based methods.

4.3 Quantitative analysis

In this study, we also compared the performance of the traditional PoCA and MLSD methods with our μ RL-enhanced approach using three quantitative metrics: peak signal-to-noise ratio (PSNR) [44], intersection over union (IoU) [45], and structural similarity index measure (SSIM). These evaluation metrics were derived from comparisons with the original images shown in Fig. 9, which were generated

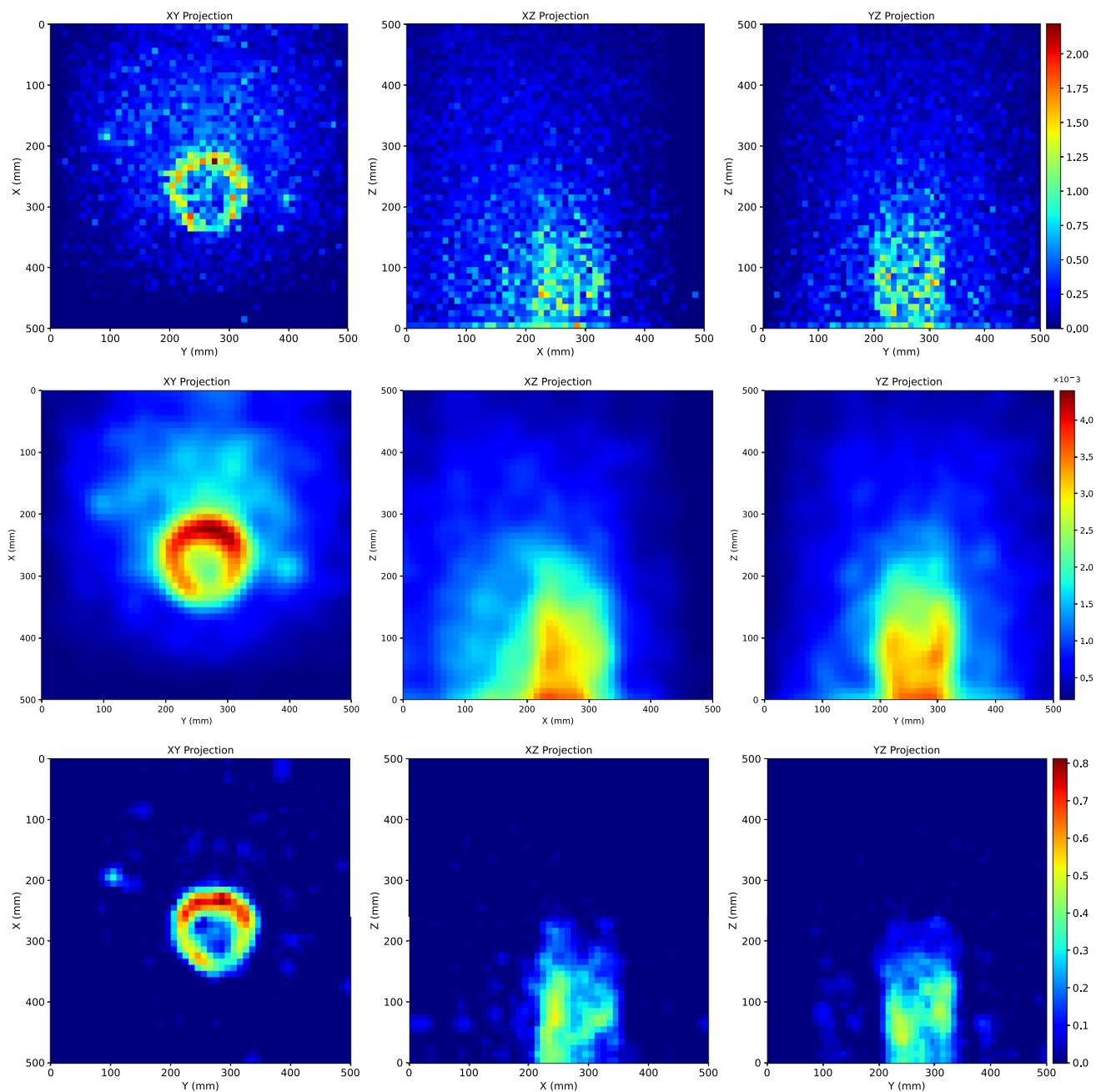


Fig. 6 (Color online) A comparison of the PoCA, MLSD method, and the μ RL-enhanced approach for the layout presented in Fig. 3c. The XY, XZ, and YZ projections generated by the PoCA method are shown in the upper panel, and the middle panel presents the corre-

sponding projections obtained using the MLSD method, whereas the lower panel illustrates the corresponding projections obtained using the μ RL-enhanced technique

using Geant4 simulations that produce position point cloud data based on geometric constructs composed of various materials.

PSNR assesses the clarity of reconstructed features by measuring the prominence of the desired signal relative to surrounding noise. The SSIM evaluates image quality from a perceptual standpoint, considering factors such as

luminance, contrast, and structural information. The IoU measures the spatial overlap between the reconstructed geometric structures and the original ones, providing a direct assessment of the reconstruction accuracy. By focusing on these metrics, we aim to provide a thorough and objective assessment of how effectively our μ RL-enhanced

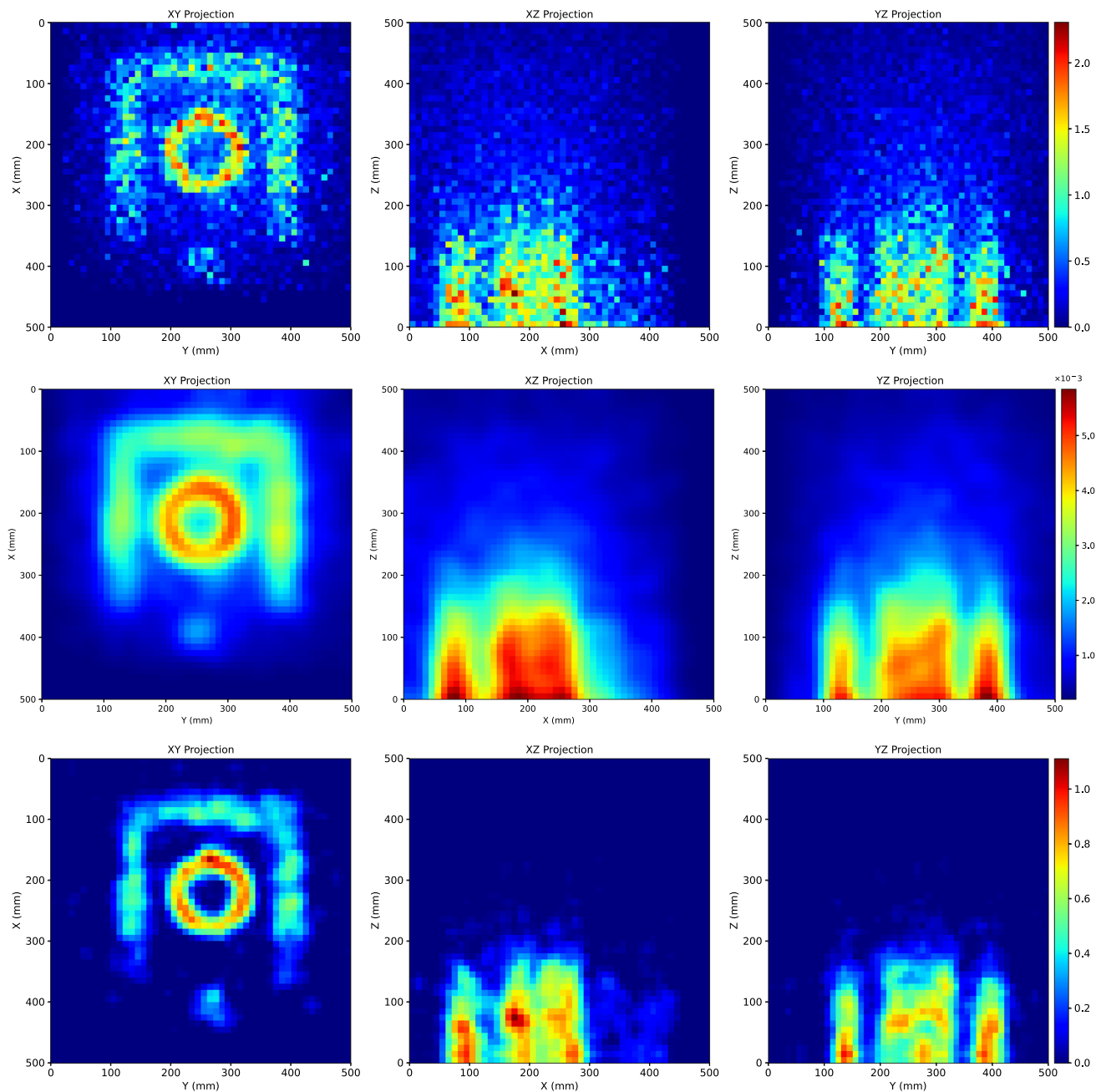


Fig. 7 (Color online) A side-by-side comparison of the PoCA, MLSD, and μ RL-enhanced methods for the layout shown in Fig. 3d. The upper, middle, and lower panels, respectively, display the XY, XZ, and YZ projections generated by the PoCA, MLSD, and μ RL-enhanced methods

pipeline reconstructs and preserves critical image details compared with the PoCA and MLSD baselines.

Table 1 shows that the μ RL-enhanced reconstructions significantly outperform the baseline PoCA algorithm, achieving average gains of 8.4% in PSNR, 307% in IoU, and 79% in SSIM. Relative to the more sophisticated MLSD method, the μ RL-enhanced attains a PSNR that is 11% higher on average, while still delivering a 9.8% increase in the IoU and a 29% boost in the SSIM. These

numbers already demonstrate a clear advantage, and inspection of representative cases further reveals the basis for this advantage. In Fig. 3a, PoCA achieves a relatively high PSNR, reflecting a low voxel-wise error, yet loses almost all object geometry, whereas MLSD restores the shape but blurs the image; in contrast, μ RL-enhanced maintains sharpness while recovering the structure. In the challenging Fig. 3d, μ RL-enhanced again surpasses both baselines, offering the clearest contours and the

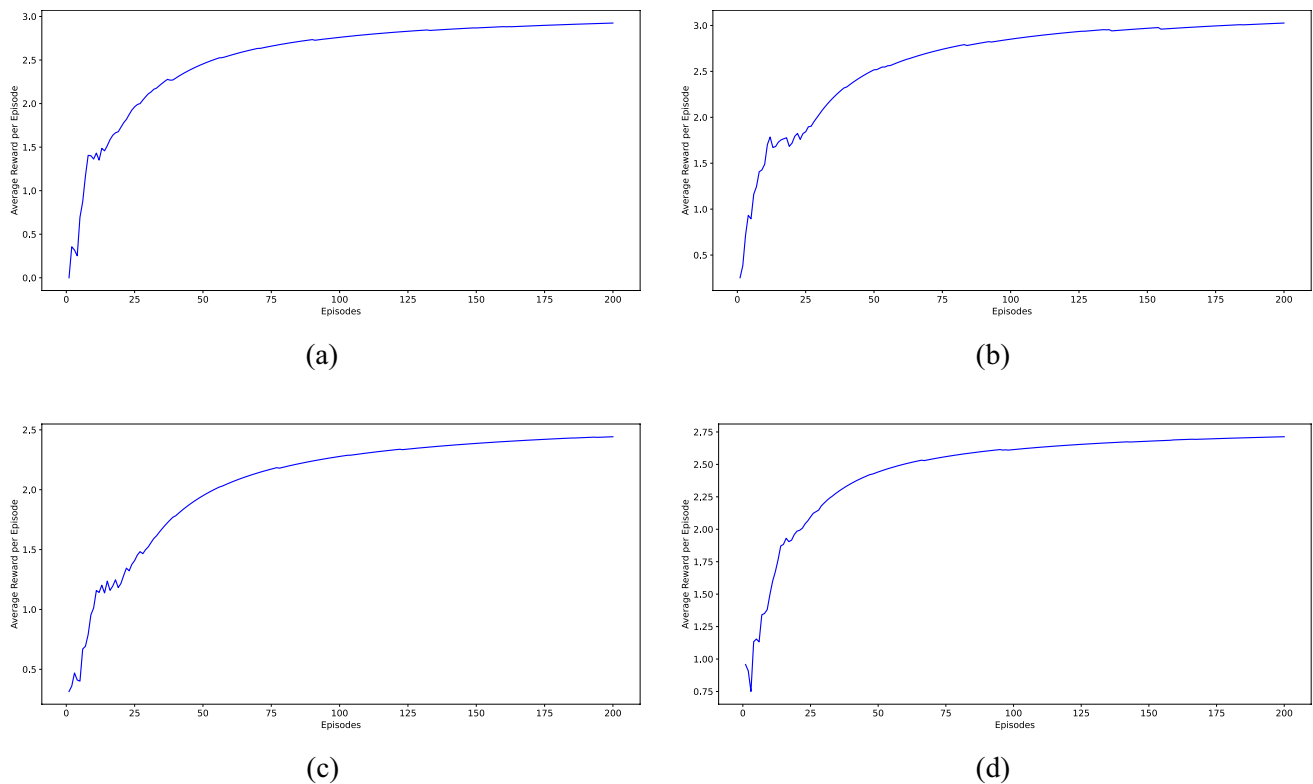


Fig. 8 (Color online) Average reward per episode during training. Panels **a** through **d** depict the average reward achieved in each episode for the corresponding layout presented in Fig. 3

most accurate overlap. Taken together, these concurrent improvements in signal fidelity, structural similarity, and overlap accuracy show that μ RL-enhanced preserves both fine details and global structure more faithfully than either reference method, thereby yielding clearer and more reliable reconstructions for downstream material identification.

5 Conclusion

This study successfully integrated RL with MST, yielding notable gains in reconstruction quality compared to the traditional PoCA and MLSD methods. By dynamically optimizing key parameters, such as the maximum scattering distance, minimum scattering angle, scattering point calculation strategy, and Gaussian smoothing, the μ RL-enhanced framework effectively addressed PoCA's reliance on static settings.

Quantitatively, the μ RL-enhanced method improves the average PSNR by 8.4%, boosts the mean IoU by 307%, and raises the SSIM by 79% relative to the PoCA baseline, yielding markedly crisper edges and more faithful structural details. When compared with the more advanced MLSD reconstruction, μ RL-enhanced method achieves an average improvement of 11% in PSNR, 9.8% in IoU, and 29% in

SSIM, confirming its ability to capture fine-scale features without compromising noise performance.

With these encouraging improvements, the current framework exhibits several limitations. First, it relies on a simplified state representation (e.g., density statistics), which restricts its ability to capture the essential spatial features. Second, the fixed step sizes in the action space limit the precision of parameter tuning. Finally, our reward function still relies on manually chosen weights based on experiments. These ad hoc weights can bias the learning process toward specific scenarios, limiting the method's effectiveness when applied to different detector setups or noise conditions. To further advance the RL framework in MST, future efforts will integrate convolutional neural networks (CNNs) for spatial-aware state encoding, adopt continuous action spaces through policy gradient methods (e.g., DDPG) for finer parameter control, and incorporate adaptive, physics-informed reward functions, leveraging scattering angle distributions to align rewards with theoretical priors, thus eliminating the need for ad hoc weights and ensuring robust generalization across diverse detector configurations and varying noise environments.

Our experimental findings highlight how feedback-driven RL methodologies can adapt MST reconstruction to varying material compositions, noise levels, and geometric

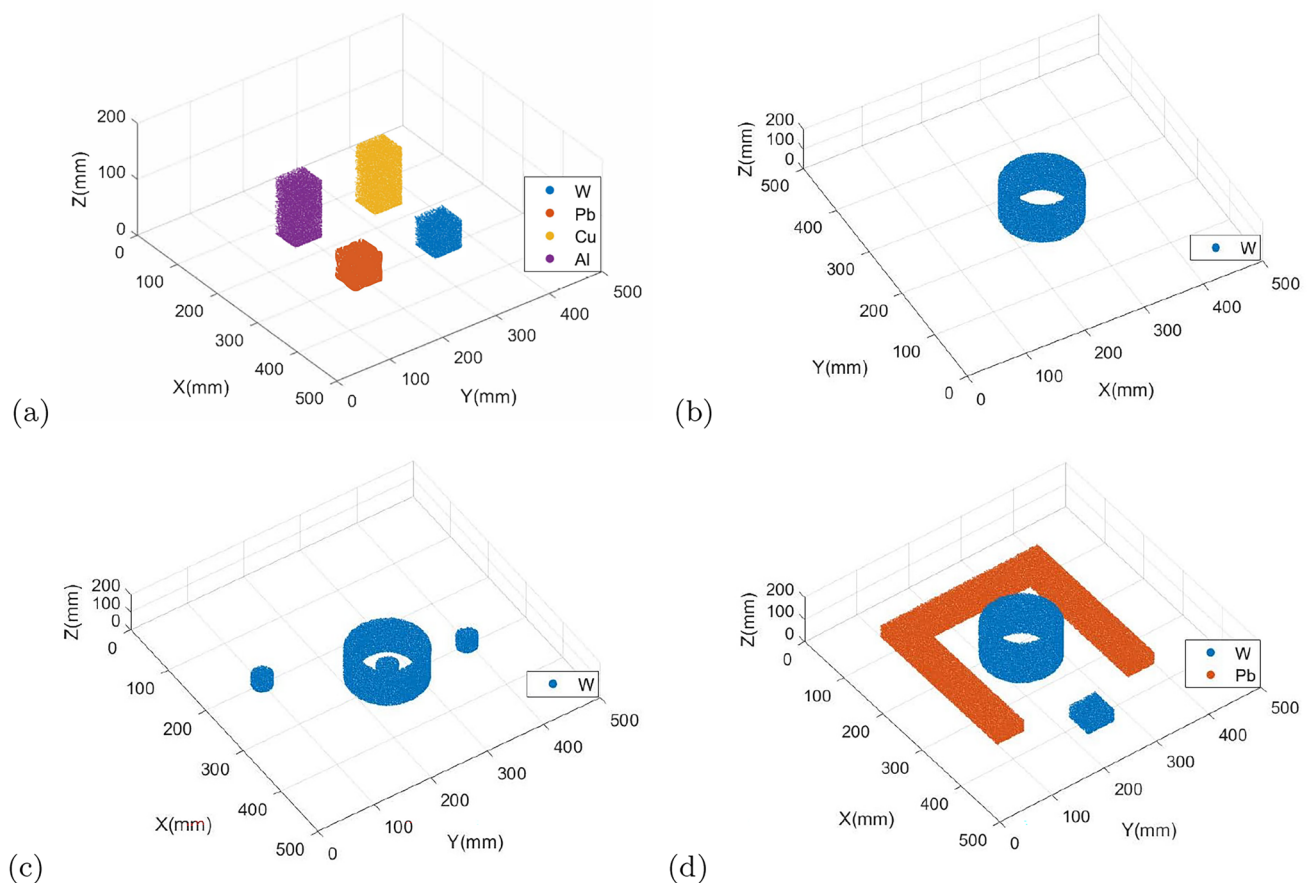


Fig. 9 (Color online) Simulated outputs as original images. Panels (a) through (d) correspond to the layouts presented in Fig. 3a–d, respectively

Table 1 Performance metrics comparison

Image	Method	PSNR (dB)	IoU	SSIM
Figure 3a	PoCA	24.29	0.06	0.50
	MLSD	19.90	0.29	0.73
	μ RL-enhanced	24.44	0.33	0.95
Figure 3b	PoCA	19.45	0.09	0.68
	MLSD	20.15	0.44	0.62
	μ RL-enhanced	21.87	0.47	0.96
Figure 3c	PoCA	18.78	0.13	0.49
	MLSD	18.85	0.42	0.77
	μ RL-enhanced	21.01	0.46	0.96
Figure 3d	PoCA	16.34	0.17	0.46
	MLSD	17.08	0.32	0.81
	μ RL-enhanced	17.76	0.35	0.87

complexities. Beyond MST, the fusion of RL and physics-driven imaging principles offers a robust paradigm for muon transmission imaging for geological exploration [46], archaeology, mineral prospecting, and other applications, where increased precision and adaptability are paramount.

By addressing these limitations, adaptive learning strategies can unlock unprecedented accuracy in next-generation muon imaging systems.

Author contributions All authors contributed to the study conception and design. Yi-Ni Wu, Yuan-Yuan Liu, Li Wang, Jian-Jie Zhang, Ning Su, Wen-Wan Ding, Xin Zhao, Zhi Zhou, Peng Zheng, and Jian-Ping Cheng contributed to material preparation, data collection, and analysis. Yi-Ni Wu wrote the first draft of the manuscript, 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.sciedb.j00186.00880> and <https://doi.org/10.57760/sciedb.j00186.00880>.

Declarations

Conflict of interest Yuan-Yuan Liu is an editorial board member for Nuclear Science and Techniques and was not involved in the editorial review, or the decision to publish this article. All authors declare that there are no conflict of interest.

References

1. W.C. Priedhorsky, K.N. Borozdin, G.E. Hogan et al., Detection of high-Z objects using multiple scattering of cosmic ray muons. *Rev. Sci. Instrum.* **74**, 4294–4297 (2003). <https://doi.org/10.1063/1.1606536>
2. C. Thomay, P. Baesso, D. Cussans et al., A novel technique to detect special nuclear material using cosmic rays. *Geosci. Instrum. Methods Data Syst.* **1**(2), 235–238 (2012). <https://doi.org/10.5194/gi-1-235-2012>
3. C. Thomay, J.J. Velthuis, P. Baesso et al., A binned clustering algorithm to detect high-Z material using cosmic muons. *J. Instrum.* **8**, P10013 (2013). <https://doi.org/10.1088/1748-0221/8/10/P10013>
4. T.B. Blackwell, The use of cosmic-rays in detecting illicit nuclear materials. University of Sheffield (2015)
5. C. Morris, C. Alexander, J.D. Bacon et al., Tomographic imaging with cosmic ray muons. *Sci. Global Sec.* **16**, 37–53 (2008). <https://doi.org/10.1080/08929880802335758>
6. L.J. Schultz, K.N. Borozdin, J.J. Gomez et al., Cosmic ray muon radiography for contraband detection. No. LA-UR-03-4804. Los Alamos National Laboratory, Los Alamos, NM (United States), 2003
7. K.N. Borozdin, G.E. Hogan, C. Morris et al., Surveillance: radiographic imaging with cosmic-ray muons. *Nature* **422**, 277 (2003). <https://doi.org/10.1038/422277a>
8. D. Poulson, J.M. Durham, E. Guardincerri et al. Cosmic ray muon computed tomography of spent nuclear fuel in dry storage casks. *Nucl. Instrum. Meth. Phys. Res. Sect. A* **842**, 48–53 (2017). <https://doi.org/10.1016/j.nima.2016.10.040>
9. S. Chatzidakis, C.K. Choi, L.H. Tsoukalas, Analysis of spent nuclear fuel imaging using multiple Coulomb scattering of cosmic muons. *IEEE Trans. Nucl. Sci.* **63**, 2866–2874 (2016). <https://doi.org/10.1109/TNS.2016.2618009>
10. S. Barnes, A. Georgadze, A. Giammanco et al., Cosmic-ray tomography for border security. *Instruments* **7**, 13 (2023). <https://doi.org/10.3390/instruments7010013>
11. S. Riggi, V. Antonuccio-Delegu, M. Bandieramonte et al., Muon tomography imaging algorithms for nuclear threat detection inside large volume containers with the Muon Portal detector. *Nucl. Instrum. Meth. Phys. Res. Sect. A* **728**, 59–68 (2013). <https://doi.org/10.1016/j.nima.2013.06.040>
12. J. Bae, S. Chatzidakis, The effect of cosmic ray muon momentum measurement for monitoring shielded special nuclear materials. *arXiv:2109.02470* (2021). <https://doi.org/10.48550/arxiv.2109.02470>
13. L.J. Schultz, Cosmic ray muon radiography. Dissertation, Portland State University, 2003
14. R.S. Sutton, A.G. Barto, Reinforcement learning: an introduction. 2nd edn. (The MIT Press, 2018)
15. L.J. Schultz, G.S. Blanpied, K.N. Borozdin et al., Statistical reconstruction for cosmic ray muon tomography. *IEEE Trans. Image Process.* **16**, 1985–1993 (2007). <https://doi.org/10.1109/TIP.2007.901239>
16. Y. Liu, Z. Zhao, Z. Chen et al., Research on the acceleration of MLS-EM reconstruction algorithm for cosmic ray muon imaging. *CT Theory and Application Research* **16**, 1–5 (2007). (in Chinese)
17. Y.Y. Liu, Z.R. Zhao, Z.Q. Chen et al., MLSD-OSEM reconstruction algorithm for cosmic ray muon radiography. *Proc. SPIE* 6913, Medical Imaging 2008: Physics of Medical Imaging, 691331 (18 March 2008). <https://doi.org/10.1117/12.769803>
18. G. Wang, L.J. Schultz, J. Qi, Bayesian image reconstruction for improving detection performance of muon tomography. *IEEE Trans. Image Process.* **18**, 1080–1089 (2009). <https://doi.org/10.1109/TIP.2009.2014423>
19. G. Wang, L. Schultz, J. Qi, Statistical image reconstruction for muon tomography using a gaussian scale mixture model. *IEEE Trans. Nucl. Sci.* **56**, 2480–2486 (2009). <https://doi.org/10.1109/TNS.2009.2023518>
20. Z.Y. Li, Z. Qian, J.H. He et al., Improvement of machine learning-based vertex reconstruction for large liquid scintillator detectors with multiple types of PMTs. *Nucl. Sci. Tech.* **33**, 93 (2022). <https://doi.org/10.1007/s41365-022-01078-y>
21. D.M. Pelt, K.J. Batenburg, J.A. Sethian, Improving tomographic reconstruction from limited data using mixed-scale dense convolutional neural networks. *J. Imaging* **4**, 128 (2018). <https://doi.org/10.3390/jimaging4110128>
22. X. Pan, Y. Zheng, Z. Zeng et al., Experimental validation of material discrimination ability of muon scattering tomography at the TUMUTY facility. *Nucl. Sci. Tech.* **30**, 120 (2019). <https://doi.org/10.1007/s41365-019-0649-4>
23. M. Ghasemi, D. Ebrahimi, Introduction to reinforcement learning. *arXiv:2408.07712* (2024). <https://doi.org/10.48550/arXiv.2408.07712>
24. M.Z. Hu, J.H. Zhang, L. Matkovic et al., Reinforcement learning in medical image analysis: concepts, applications, challenges, and future directions. *J. Appl. Clin. Med. Phys.* **24**, e13898 (2023). <https://doi.org/10.1002/acm2.13898>
25. S.K. Zhou, H.N. Le, K. Luu et al., Deep reinforcement learning in medical imaging: a literature review. *Med. Image Anal.* **73**, 102193 (2021). <https://doi.org/10.1016/j.media.2021.102193>
26. A.V. Bernstein, E.V. Burnaev, O.N. Kachan, Reinforcement learning for computer vision and robot navigation. In: Perner, P. (eds) *Machine Learning and Data Mining in Pattern Recognition. MLDM 2018. Lecture Notes in Computer Science*, vol. 10935. Springer, Cham, 2018. https://doi.org/10.1007/978-3-319-96133-0_20
27. N. Le, V.S. Rathour, K. Yamazaki et al., Deep reinforcement learning in computer vision: a comprehensive survey. *Artif. Intell. Rev.* **55**, 2733–2819 (2022). <https://doi.org/10.1007/s10462-021-10061-90>
28. G.W. Taylor, A reinforcement learning framework for parameter control in computer vision applications. *First Canadian Conference on Computer and Robot Vision*, 2004. Proceedings., London, ON, Canada, 2004, pp. 496–503. <https://doi.org/10.1109/CCCRV.2004.1301489>
29. A. Alansary, O. Oktay, Y. Li et al., Evaluating reinforcement learning agents for anatomical landmark detection. *Med. Image Anal.* **53**, 156–164 (2019). <https://doi.org/10.1016/j.media.2019.02.007>
30. A. Vlontzos, A. Alansary, K. Kamnitsas et al., Multiple landmark detection using multi-agent reinforcement learning. *arXiv:1907.00318* (2019). <https://doi.org/10.48550/arxiv.1907.00318>
31. R.L. Smith, I.M. Ackerley, K. Wells et al., Reinforcement learning for object detection in PET imaging. *IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*. Manchester, UK **2019**, 1–4 (2019). <https://doi.org/10.1109/NSS/MIC42101.2019.9060031>
32. F. Sahba, H.R. Tizhoosh, M.M.A. Salama, A reinforcement learning framework for medical image segmentation. *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, Vancouver, BC, Canada, pp. 511–517 (2006). <https://doi.org/10.1109/IJCNN.2006.246725>
33. Z.Q. Tian, X.Y. Si, Y.Y. Zheng et al., Multi-step medical image segmentation based on reinforcement learning. *J. Ambient. Intell. Human. Comput.* **13**, 5011–5022 (2022). <https://doi.org/10.1007/s12652-020-01905-3>
34. S.H. Sun, J. Hu, M.Q. Yao et al., Robust multimodal image registration using deep recurrent reinforcement learning. In:

- Jawahar, C., Li, H., Mori, G., Schindler, K. (eds) Computer Vision – ACCV 2018. ACCV 2018. Lecture Notes in Computer Science, vol 11362. Springer, Cham. https://doi.org/10.1007/978-3-030-20890-5_33
35. J. Hu, Z. Luo, X. Wang et al., End-to-end multimodal image registration via reinforcement learning. *Med. Image Anal.* **68**, 101878 (2021). <https://doi.org/10.1016/j.media.2020.101878>
 36. C. Shen, Y. Gonzalez, L. Chen et al., Intelligent parameter tuning in optimization-based iterative CT reconstruction via deep reinforcement learning. *IEEE T. Med. Imaging.* **37**, 1430 (2018). <https://doi.org/10.1109/TMI.2018.2823679>
 37. L. Paletta, A. Pinz, Active object recognition by view integration and reinforcement learning. *Robot. Auton. Syst.* **31**, 71–86 (2000). [https://doi.org/10.1016/S0921-8890\(99\)00079-2](https://doi.org/10.1016/S0921-8890(99)00079-2)
 38. R. Furuta, N. Inoue, T. Yamasaki, PixelRL: fully convolutional network with reinforcement learning for image processing. *IEEE T. Multimedia.* **22**, 1704–1719 (2019). <https://doi.org/10.1109/TMM.2019.2960636>
 39. J. Zhang, Q. Zhang, X. Zhao et al., Boosting denoisers with reinforcement learning for image restoration. *Soft. Comput.* **26**, 3261–3272 (2022). <https://doi.org/10.1007/s00500-022-06840-3>
 40. J. Huang, J. Su, Q. Chang, Graph neural network and multi-agent reinforcement learning for machine-process-system integrated control to optimize production yield. *J. Manuf. Syst.* **64**, 81–93 (2022). <https://doi.org/10.1016/j.jmsy.2022.05.018>
 41. P. Yu, Z. Pan, Z. He et al., A new efficient imaging reconstruction method for muon scattering tomography. *Nucl. Instrum. Meth. Phys. Res. Sect. A* **1069**, 169932 (2024). <https://doi.org/10.1016/j.nima.2024.169932>
 42. Q.G. Wen, Research on rapid imaging with cosmic ray muon scattering tomography. *Sci. Rep. UK* **13**, 19718 (2023). <https://doi.org/10.1038/s41598-023-47023-w>
 43. C.J.C.H. Watkins, P. Dayan, Q-learning. *Mach. Learn.* **8**, 279–292 (1992). <https://doi.org/10.1007/BF00992698>
 44. S. Luo, Y. Huang, X. Ji et al., Hybrid model for muon tomography and quantitative analysis of image quality. *Nucl. Sci. Tech.* **33**, 81 (2022). <https://doi.org/10.1007/s41365-022-01070-6>
 45. K. Hartling, F. Mahoney, E.T. Rand et al., A comparison of algebraic reconstruction techniques for a single-detector muon computed tomography system. *Nucl. Instrum. Meth. Phys. Res. Sect. A* **987**, 164834 (2021). <https://doi.org/10.1016/j.nima.2020.164834>
 46. Y. Cheng, R. Han, Z. Li et al., Imaging internal density structure of the Laoheishan volcanic cone with cosmic ray muon radiography. *Nucl. Sci. Tech.* **33**, 88 (2022). <https://doi.org/10.1007/s41365-022-01072-4>

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.