

A deep unsupervised clustering-based post-processing framework for high-fidelity Cerenkov luminescence tomography

Cite as: J. Appl. Phys. **128**, 193104 (2020); <https://doi.org/10.1063/5.0025877>

Submitted: 19 August 2020 . Accepted: 09 November 2020 . Published Online: 20 November 2020

 Xin Cao, Jun Zhang, Jianan Yang, Chunxiao Fan, Fengjun Zhao,  Wei Zhou, Lin Wang, Guohua Geng, Mingquan Zhou, and Xueli Chen



View Online



Export Citation



CrossMark

ARTICLES YOU MAY BE INTERESTED IN

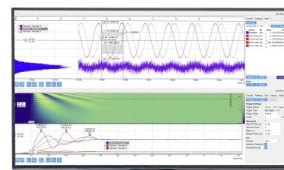
[The relationship between the doping concentration and \$d^0\$ ferromagnetism in n-type 4H-SiC](#)
Journal of Applied Physics **128**, 193901 (2020); <https://doi.org/10.1063/5.0028458>

[Photothermal spectroscopy: A promising tool for nanofluids](#)
Journal of Applied Physics **128**, 190901 (2020); <https://doi.org/10.1063/5.0024332>

[Evolution of the conductive filament with cycling in TaO_x-based resistive switching devices](#)
Journal of Applied Physics **128**, 194501 (2020); <https://doi.org/10.1063/5.0032494>

Challenge us.

What are your needs for
periodic signal detection?



Zurich
Instruments



A deep unsupervised clustering-based post-processing framework for high-fidelity Cerenkov luminescence tomography

Cite as: J. Appl. Phys. 128, 193104 (2020); doi: 10.1063/5.0025877

Submitted: 19 August 2020 · Accepted: 9 November 2020 ·

Published Online: 20 November 2020



Xin Cao,¹  Jun Zhang,¹ Jianan Yang,¹ Chunxiao Fan,² Fengjun Zhao,¹ Wei Zhou,^{1,a)}  Lin Wang,¹ Guohua Geng,¹ Mingquan Zhou,³ and Xueli Chen^{4,a)}

AFFILIATIONS

¹School of Information Sciences and Technology, Northwest University, Xi'an, Shaanxi 710127, China

²School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, Anhui 230601, China

³Engineering Research Center of Virtual Reality and Applications, Ministry of Education, Beijing Key Laboratory of Digital Preservation and Virtual Reality for Cultural Heritage, Beijing Normal University, Beijing, China

⁴Engineering Research Center of Molecular and Neuro Imaging of Ministry of Education & School of Life Science and Technology, Xidian University, Xi'an, Shaanxi 710071, China

^{a)}Authors to whom correspondence should be addressed: mczhouwei12@gmail.com and xlchen@xidian.edu.cn

ABSTRACT

Cerenkov Luminescence Tomography (CLT) is a promising optical molecular imaging technology. It involves the three-dimensional reconstruction of the distribution of radionuclide probes inside a single object to indicate a tumor's localization and distribution. However, reconstruction using CLT suffers from severe ill-posedness, resulting in numerous artifacts within the reconstructed images. These artifacts influence the visual effect and may misguide the medical professional (diagnostician), resulting in a wrong diagnosis. Here, we proposed a deep unsupervised clustering-based post-processing framework to eliminate artifacts and facilitate high-fidelity CLT. First, an initial reconstructed image was obtained by a specific reconstruction method. Second, voxel data were generated based on the initial reconstructed result. Third, these voxels were divided into three groups, and only the group with the highest mean intensity was chosen as the final reconstructed result. A group of numerical simulation and *in vivo* mouse-based experiments were conducted to assess the presented framework's feasibility and potential. The results indicated that the proposed framework could reduce the number of artifacts effectively. The reconstructed image's shape and distribution were more similar to the actual light source than those obtained without the proposed framework.

Published under license by AIP Publishing. <https://doi.org/10.1063/5.0025877>

I. INTRODUCTION

Cerenkov luminescence imaging (CLI) is a promising imaging modality for optical molecular imaging. CLI has captured the attention of many researchers in recent years, and its advantages include low cost, widely available clinical radionuclide probes, and a relatively high throughput.^{1–3} Briefly, CLI is an optical imaging technique that can identify the distribution of radionuclide probes (in two dimensions) by collecting the Cerenkov Luminescence (CL) emitted from these probes as the isotope decays. Since the first application of CLI to detect the ²⁻¹⁸fluoro-D-glucose (¹⁸F-FDG) inside one mouse by Robertson *et al.*, CLI has been successfully used in the early diagnosis of tumors, the treatment of various

diseases, and in clinical surgery.^{4–8} One method to improve CLI's capability is to obtain a three-dimensional (3D) distribution of radionuclide probes. To this end, Li *et al.* proposed a tomographic method named Cerenkov Luminescence Tomography (CLT) in 2010.⁹ Subsequently, many researchers have devoted themselves to developing methods to improve the accuracy of CLT reconstruction. These include the *prior* information- and shrinking permission-based methods.^{10–16} It is worth noting that most of these reconstruction methods are based on the finite element method (FEM). Using FEM, the whole domain of the imaging object is discretized into a set of nodes, and this reconstruction aims to determine the intensity of these nodes. The resolution of these methods, however,

depends heavily on the discretization density of finite element meshes. Over-dense grids will make reconstruction impossible, and sparse grids will result in low resolution of reconstructed images.

Conversely, similar to traditional optical tomography, CLT's system matrix is usually ill-conditioned, leading to a highly ill-posed reconstruction problem.^{5,17–19} Many methods have been proposed for CLT to reduce ill-posedness significantly, but these methods have been unable to produce extremely accurate results. In the reconstructed image, nodes surrounding the target object show a relatively high reconstruction intensity. These nodes are considered artifacts and would mislead the users to make a wrong diagnostic decision. Elimination of these artifacts requires an efficient post-processing method. Unfortunately, up until now, only a few methods have been proposed. One commonly used method is the artificial threshold (ArT) method, which filters these artifacts by setting a fixed threshold. Another method is the adaptive threshold setting method proposed by Yi *et al.*²⁰ Here, they sorted the nodes in the reconstructed image and calculated a threshold to filter them. In this way, the adaptive threshold setting method can avoid the human intervention required in the ArT method. However, these methods have two major drawbacks. First, both of them are single threshold-based methods. If nodes surrounding or far-from the target object have a significant intensity, it is hard to determine a suitable threshold to filter those nodes out. Second, these single threshold-based methods may reduce the quality of the reconstructed image. If some nodes having a relatively low intensity exist in the reconstructed target volume, these methods will also filter out these nodes and reduce the reconstructed image's fidelity.

Here, inspired by a deep neural network successfully applied in imaging processing,^{21,22} we proposed a deep unsupervised clustering-based post-processing framework for CLT. This framework can improve the resolution and reduce the artifacts of the reconstructed image, facilitating high-fidelity CLT. First, the initial reconstructed results were obtained using a CLT reconstruction method. Second, the 3D space where the reconstructed image resides was transformed into voxelized data. These high density grid-based voxelized data support increasing the spatial resolution of the reconstructed image. Third, the deep unsupervised clustering algorithm called deep embedded clustering (DEC),²³ which jointly learns feature representation and cluster assignments, was trained to cluster these voxels into three groups. Finally, the group with the highest mean intensity was selected as the final reconstructed result. Compared with the traditional post-processing methods mentioned above, our proposed framework has two main advantages. These advantages include (1) the spatial resolution of the reconstructed result can be significantly increased after voxelization, which will be beneficial for obtaining a relatively precise shape and providing high fidelity of the unknown target; (2) the proposed framework can achieve a robust result, preventing the interference of the noise voxels with a relatively high intensity, which cannot be filtered out by merely setting a threshold.

The rest of this paper is structured as follows. The ArT post-processing method, DEC algorithm, and post-processing framework based on DEC for CLT are presented in Sec. II. Several numerical simulation experiments and *in vivo* mouse experiments are described in Sec. III to demonstrate the proposed method's effectiveness. Finally, we provide a conclusion in Sec. IV.

II. METHODS

A. Artificial threshold (ArT) method

After obtaining the initial reconstructed results, a traditional method to eliminate the artifacts is to set a threshold using an empirical intensity to filter out these artifacts. This procedure can be defined as

$$r_i = \begin{cases} r_i, & r_i > \gamma \cdot \max(r), \\ 0, & \text{otherwise}, \end{cases} \quad (1)$$

where r denotes all the intensities of the reconstructed nodes and r_i denotes the i th node's intensity. γ is an empirical number with a range of 0–1.

B. Deep embedded clustering (DEC) algorithm

Deep embedded clustering (DEC) is an unsupervised clustering method based on the deep learning framework and can achieve excellent performance for clustering studies. Briefly, DEC can be divided into three stages. First, the Stacked Denoising Autoencoder (SDAE), a famous deep neural network with a powerful expressive ability, is trained using original data. The trained SDAE can map multi-dimensional data into low-dimensional embedded vector space, which can be defined as

$$f_{\theta}: X \rightarrow E, \quad (2)$$

where X denotes the original data space. E denotes an embedded vector space learned by SDAE, in which the embedded vector is an effective feature representation of the original input data sample. θ denotes the set of all the parameters of SDAE. Using the middle layer of SDAE as the output layer, we can pass the original data through the trained SDAE to get embedded data points. Then, k -means clustering is performed in the feature space E to obtain k cluster center C .

Second, the Student's t -distribution is used as a kernel to measure the similarity between the embedded vector e_i and clustering center c_j as

$$q_{ij} = \frac{(1 + \|e_i - c_j\|^2/v)^{-\frac{v+1}{2}}}{\sum_j (1 + \|e_i - c_j\|^2/v)^{-\frac{v+1}{2}}}, \quad (3)$$

where v represents the degree of freedom, which is used for cross-validation. Here, v is set to 1 because we cannot cross-validate v on a validation set in the unsupervised setting. e_i corresponds to the original data x_i after embedding. c_j denotes the center of the j th cluster in the embedded space, and $\|\cdot\|^2$ denotes L2-norm. So, q_{ij} can be interpreted as the probability of assigning data i to cluster j .

Finally, the loss function of DEC is defined based on Kullback–Leibler (KL) divergence as

$$L = KL(P||Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}, \quad (4)$$

where P denotes the target distribution, and it is defined as

$$p_{ij} = \frac{q_{ij}^2 / \sum_i q_{ij}}{\sum_j q_{ij}^2 / \sum_i q_{ij}}. \tag{5}$$

We jointly optimize the cluster centers $\{c_j\}$ and deep neural network (DNN) parameters θ using Stochastic Gradient Descent (SGD) with momentum. The gradients of L with respect to feature-space embedding of each data point e_i and each cluster centroid c_j are computed as

$$\frac{\partial L}{\partial e_i} = \frac{v+1}{v} \sum_j \left(1 + \frac{\|e_i - c_j\|^2}{v}\right)^{-1} \times (p_{ij} - q_{ij})(e_i - c_j), \tag{6}$$

$$\frac{\partial L}{\partial c_j} = -\frac{v+1}{v} \sum_i \left(1 + \frac{\|e_i - c_j\|^2}{v}\right)^{-1} \times (p_{ij} - q_{ij})(e_i - c_j). \tag{7}$$

The gradients $\frac{\partial L}{\partial e_i}$ are then passed down to the DNN and used in standard backpropagation to compute the DNN's parameter gradient $\frac{\partial L}{\partial \theta}$. We stop our procedure for discovering cluster assignments when less than *tol*% of points change cluster assignment between two consecutive iterations, similar to a previous study.²³ Notably, in our work, we implemented the DEC network using PyTorch (Python 3.6), with the hardware environment (3.6 GHz Intel Core i9-9900K CPU, 128 GB RAM, and four GTX 2080Ti GPUs).

C. DEC-based post-processing framework for CLT

As mentioned in Sec. I, the original reconstructed image includes artifacts. In this section, a post-processing framework for CLT is proposed. The flow chart of the proposed framework is shown in Fig. 1.

First, based on the collected CL signals, a whole region reconstruction is carried out to obtain the reconstructed result. It is worth noting that the reconstructed result is a set of nodes, and each of them has a reconstructed intensity. The result can be defined as

$$R = \{v_i | v_i \in T_j, val(v_i) \geq 0, i = 1, 2, \dots, n, j = 1, 2, \dots, m\}, \tag{8}$$

where v_i denotes the i th node, T_j is j th tetrahedron which contains four nodes, and $val(v_i)$ is the intensity of v_i .

As the distribution of the nodes is not uniform, the number of the nodes in the region of the target region may be small, which will lower the reconstructed image's resolution. Hence, we voxelized the space where the reconstructed result resides to increase the number of nodes significantly, which improved the spatial resolution. Assuming the maximum and minimum intensity of each axis of R is $A = \{x_{\min}, x_{\max}; y_{\min}, y_{\max}; z_{\min}, z_{\max}\}$, a minimum cube of the reconstructed domain can be constructed, with the size of $(x_{\max} - x_{\min}) \times (y_{\max} - y_{\min}) \times (z_{\max} - z_{\min})$. Then, the cube is divided into small cubes, and the reconstructed result is transformed into a set of voxels. The intensity of each voxel is

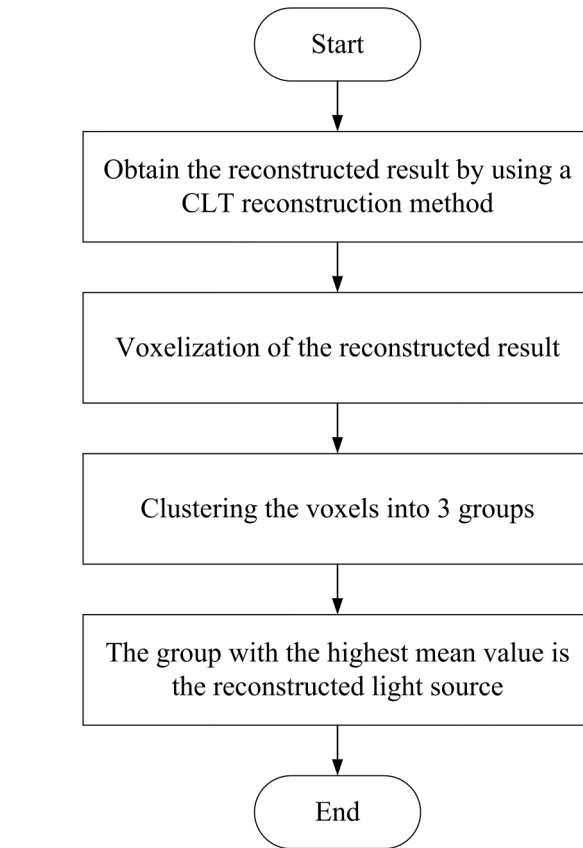


FIG. 1. Flow chart of the proposed post-processing method.

calculated as

$$val(p_i) = value(a) \times \alpha + value(b) \times \beta + value(c) \times \gamma + value(d) \times \delta, \tag{9}$$

where p_i denotes the i th voxel. It is worth noting that p_i is inside one tetrahedron, which has four vertices, and each vertex has a reconstructed intensity. Here, $a, b, c,$ and d denote the four vertices and $value(a), value(b), value(c),$ and $value(d)$ denote their intensities. As to the definition of α , it is defined as

$$\alpha = \frac{D(p_i, F_{bcd})}{D(a, F_{bcd})}, \tag{10}$$

where F_{bcd} denotes the plane formed by points $b, c,$ and d . $D(p_i, F_{bcd})$ denotes the distance between the center of p_i and F_{bcd} , and $D(a, F_{bcd})$ denotes the distance between a and F_{bcd} . Similarly, $\beta,$

γ , and δ can be defined as

$$\beta = \frac{D(P, F_{acd})}{D(b, F_{acd})}, \gamma = \frac{D(P, F_{abd})}{D(c, F_{abd})}, \delta = \frac{D(P, F_{abc})}{D(d, F_{abc})}. \quad (11)$$

Then, the voxels with intensity larger than 0 are clustered into three groups using DEC. We first construct a series of 3D patches. Each patch's size is $5 \times 5 \times 5$, and the center voxel in each patch is a voxel with its intensity larger than 0, and each patch is then converted into a vector. In this work, the SDAE bottleneck is set to 10, while the parameter k used in k -means is set to 3. Then, the average intensity of each group is calculated,

$$I_{avg} = \frac{\sum_{j=1}^m val(p_j)}{m}, \quad (12)$$

where m denotes the number of the voxels in one group and $val(p_j)$ denotes the intensity of one voxel in the group.

Finally, the group with the highest mean intensity is chosen as the reconstructed target.

III. EXPERIMENTS AND RESULTS

A. Experimental setup

A series of numerical simulations and *in vivo* mouse-based experiments were designed and conducted to evaluate the proposed framework's performance. To ensure the proposed framework's broad applicability, four different reconstruction algorithms were developed in each experiment. These include the filtered maximum likelihood expectation maximization (fMLEM),²⁴ the conjugate gradient least squares (CGLS),²⁵ the gradient projection for sparse reconstruction (GPSR),²⁶ and the incomplete variables truncated conjugate gradient (IVTCG).²⁷ The original reconstructed results were then transformed into a set of voxels using the method described in Sec. II C. Notably, in this paper, the voxel's side length was set to 0.1 mm, which significantly increased the number of elements with an intensity greater than 0. The experimental results were then quantitatively analyzed using the volume ratio (VR), the distance between the center of mass (DCoM), the ratio between the interacted volume and total volume (IVTVR), structural similarity index (SSIM), and the contrast-to-noise ratio (CNR).

The VR is defined as the ratio of the volume of the reconstructed area to the real CL source area,

$$VR = \frac{\min(V_{Real}, V_{Rec})}{\max(V_{Real}, V_{Rec})}, \quad (13)$$

where V_{Real} denotes the volume of the real CL source, while V_{Rec} denotes the volume of the reconstructed area. It should be noted that we select $\max(V_{Real}, V_{Rec})$ as the denominator and $\min(V_{Real}, V_{Rec})$ as the numerator to constrain the intensity to the range of [0, 1].

The DCoM is the distance between the center of mass (CoM) of the reconstructed area and that of the real CL source area,

$$DCoM = |C_{Rec} - C_{Real}|, \quad (14)$$

where C_{Rec} denotes the center coordinate of the reconstructed area, while C_{Real} denotes the center coordinate of the real CL source area.

IVTVR is defined in Ref. 28 as

$$IVTVR = \frac{\Omega_{Rec} \cap \Omega_{Real}}{\Omega_{Rec}}, \quad (15)$$

where Ω_{Rec} denotes the region containing reconstructed voxels, while Ω_{Real} denotes the region containing CL source.

SSIM is the index used to assess the structural similarity between two objects. Here, the definition is shown as

$$SSIM(\Psi_{Real}, \Psi_{Rec}) = \frac{(2\mu_1\mu_2 + c_1)(2\sigma_{1,2} + c_2)}{(\mu_1^2 + \mu_2^2 + c_1)(\sigma_1^2 + \sigma_2^2 + c_2)}, \quad (16)$$

where Ψ denotes the region of $\Omega_{Rec} \cup \Omega_{Real}$, μ_1 denotes the mean intensity of Ψ_{Real} , and μ_2 denotes the mean intensity of Ψ_{Rec} ; σ_1^2 is the variance of Ψ_{Real} , while σ_2^2 denotes the variance of Ψ_{Rec} ; $\sigma_{1,2}$ is the square root of the covariance between Ψ_{Real} and Ψ_{Rec} ; $c_1 = (k_1L)^2$ and $c_2 = (k_2L)^2$ are two variables to stabilize the division with the weak denominator, where L is the dynamic range of the voxel intensity (the intensity is 1 in this study), $k_1 = 0.01$ and $k_2 = 0.03$. It should be noted that a large intensity of SSIM usually corresponds to a high degree of similarity between the two objects.

CNR is defined as

$$CNR = \frac{\mu_{ROI} - \mu_{BCK}}{(m_{ROI}\sigma_{ROI}^2 + m_{BCK}\sigma_{BCK}^2)^{1/2}}, \quad (17)$$

where μ_{ROI} and μ_{BCK} are the mean values in the ROI and background, respectively; σ_{ROI}^2 and σ_{BCK}^2 are the variances; and m_{ROI} and m_{BCK} are the number of nodes included in the ROI and background, respectively. It should be noted that the ROI is where the light source is located, while the region of background selected here is a region far away from the light source.

B. Numerical simulation-based experiments

A heterogeneous cylindrical model was designed to carry out the numerical simulation-based experiments. The cylinder has a radius of 10 mm and a height of 30 mm. It is composed of five components representing five different tissues. According to Refs. 29 and 30, the optical parameters of all components are presented in Table I. Two experiments were conducted to investigate the performance of the proposed framework thoroughly. The first

TABLE I. Optical parameters used in the numerical simulation experiments.

Tissue	μ_a/mm^{-1}	μ_s/mm^{-1}	g
T1	0.016	0.510	0.9
T2	0.021	2.864	0.9
T3	0.011	1.053	0.86
T4	0.036	2.246	0.9
T5	0.012	2.472	0.9

used a single light source and the second a double light source. In both experiments, the cylindrical model was divided into 6058 nodes and 31 591 internal tetrahedral finite element meshes using COMSOL Multiphysics V5.2 (COMSOL Inc., USA) software. The surface photon energy distribution was generated using the Monte Carlo (MC) method implemented using the Molecular Optical Simulation Environment (MOSE) software.³¹

In the first experiment, the light source is a sphere with a radius of 0.5 mm. The center is set at (0.0, 6.0, 15.0) mm. The power density is set as 1.0 nW/mm^3 . Figure 2(a) shows the physical model of the single light source-based experiment, while Fig. 2(b) is the cross-sectional view of the model. It should be noted that the parameter γ in the ArT method is a fixed intensity between 0 and 1, which is relatively difficult to determine accurately. Figure 3

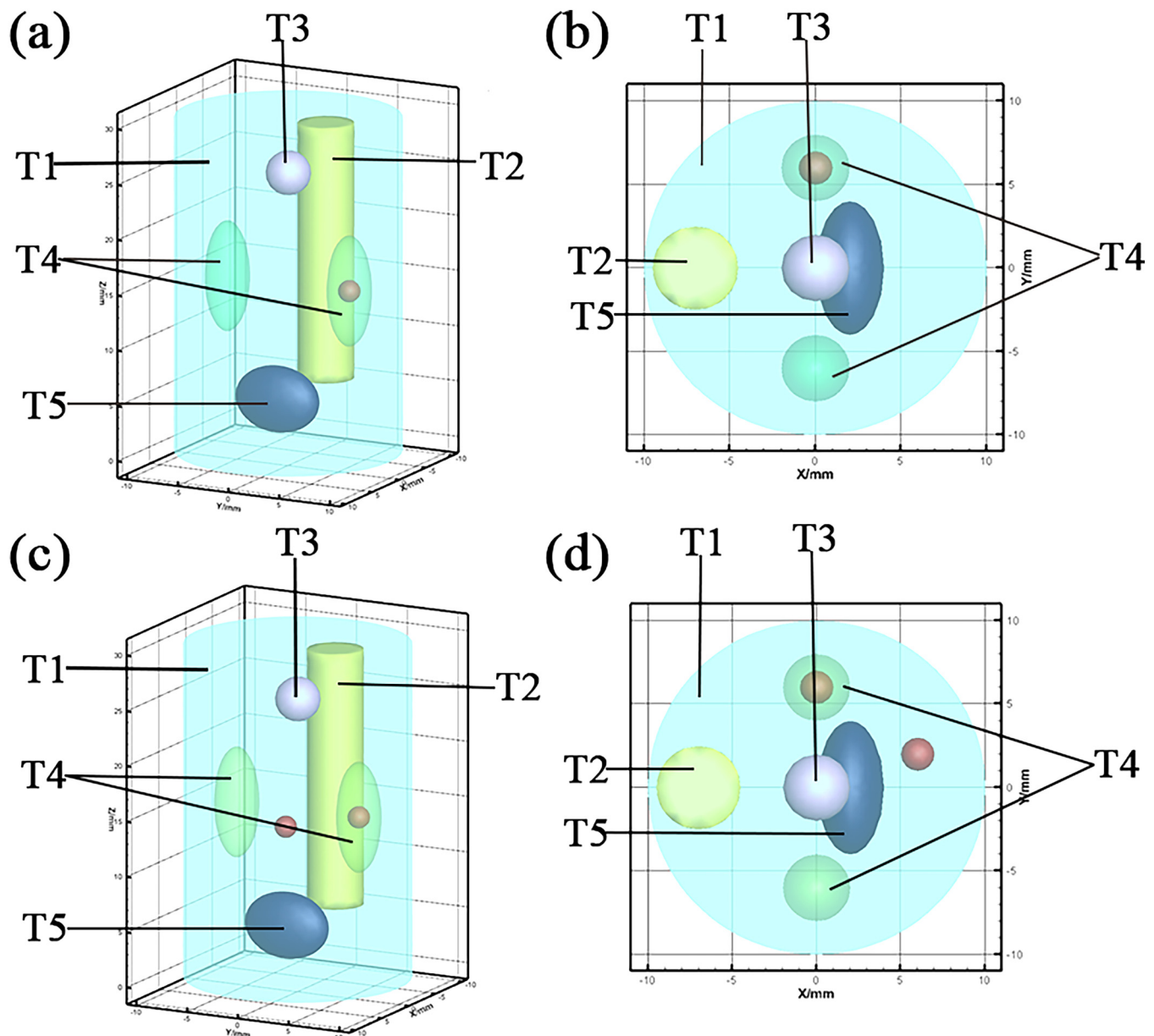


FIG. 2. Heterogeneous numerical phantom-based physical models. (a) is the single-source-based physical model, (c) is the double-source-based physical model. (b) and (d) are the two-dimensional cross-sectional views of (a) and (c).

shows the result reconstructed by IVTCG, where Fig. 3(a) is the original result and Figs. 3(b)–3(d) show the results of the ArT method with γ of 30%, 70%, and 90%, respectively. A small intensity will result in many artifacts, while a large intensity will greatly shrink the reconstructed area, resulting in poor fidelity. In the following experiments, we chose the γ intensity as 70%, which is also an empirical intensity used in the previous work.³²

The reconstructed and the post-processing results by different post-processing methods are shown in Fig. 4. Figure 4(a1) presents the original result reconstructed by the IVTCG method, and Figs. 4(b1) and 4(c1) are the post-processing results obtained by the ArT method and the DEC-based post-processing framework. Figures 4(a2)–4(c2) show the originally reconstructed result by the GPSR method, the post-processing results obtained by the ArT method, and the DEC-based post-processing framework, respectively. Figures 4(a3)–4(c3) are the corresponding results of the fMLEM method, and Figs. 4(a4)–4(c4) detail the results of the CGLS method. The red circles labeled the position of the actual light source. From Fig. 4, we observe that the original results contain many artifacts, which can be misleading to users. However, the post-processing results obtained using the ArT method had a large amount of lost information, resulting in low fidelity. The DEC-based post-processing framework worked well in all the observed cases, whose results fit well with the actual light source

distribution. Other indicators were calculated and demonstrated in Fig. 5. We observed that our proposed framework effectively filtered out the artifacts, and the shape of the reconstructed result is more similar to the real light source. It is worth noting, from the IVTVR, that the artificial threshold (ArT) method has a higher intensity than that obtained by our framework. This is because the voxel with the largest intensity is inside the real light source area, and a relatively large threshold can filter out other voxels whose intensity is smaller than 70% of the largest intensity.

Next, a double-source-based experiment was carried out to demonstrate the proposed framework's applicability to situations with multiple light sources. Figures 2(c) and 2(d) depict the double-source-based experiment's physical model. Figure 6 shows the reconstructed and post-processing results using different post-processing methods for the double-source-based experiment. Figure 6(a1) presents the original result reconstructed by the IVTCG method, and Figs. 6(b1) and 6(c1) are the post-processing results obtained by the ArT method and the DEC-based post-processing framework. Figures 6(a2)–6(c2) show the originally reconstructed result by the GPSR method, the post-processing results obtained by the ArT method, and the DEC-based post-processing framework, respectively. Figures 6(a3)–6(c3) are the corresponding results of the fMLEM method, and Figs. 6(a4)–6(c4) detail the results of the CGLS method. Red circles indicated the

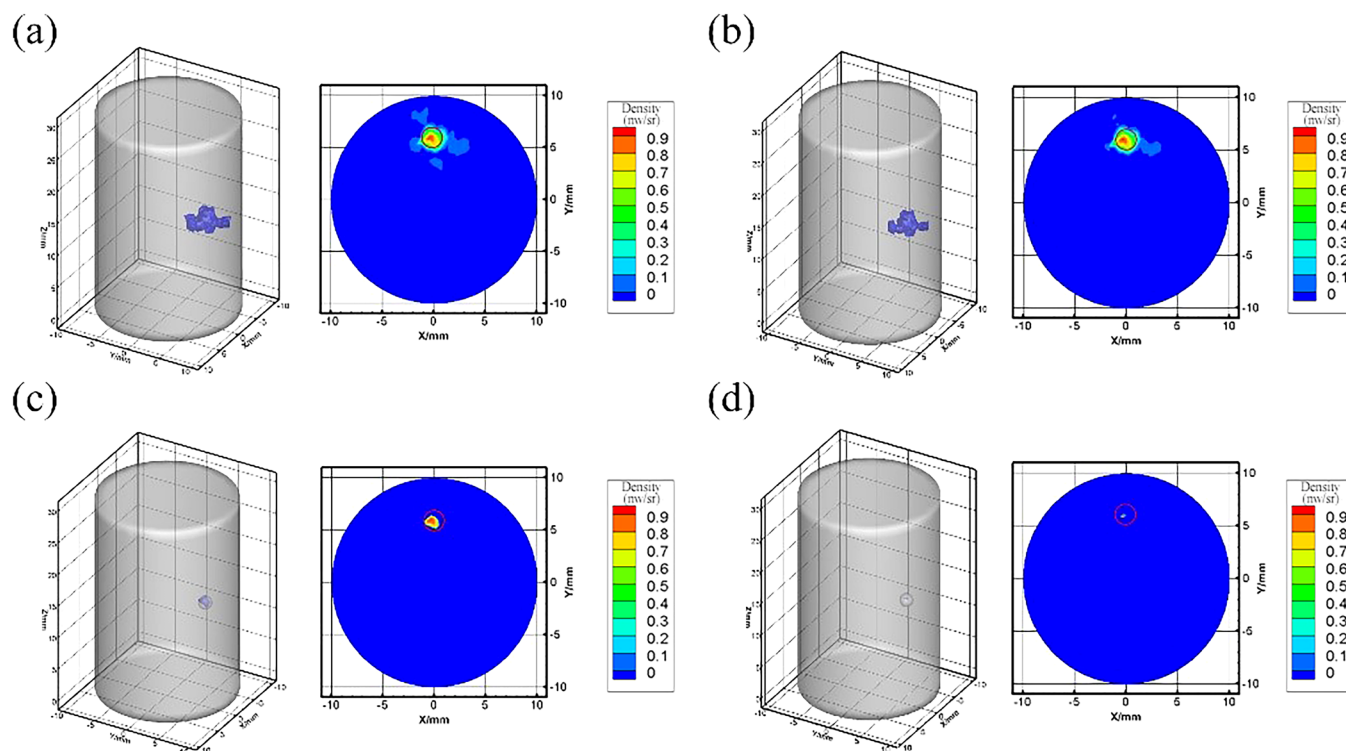


FIG. 3. Investigation of the effect of γ intensity on the final result. (a) The original result is reconstructed by the IVTCG method, while (b), (c), and (d) are the post-processing results by the ArT method with γ intensities of 30%, 70%, and 90%, respectively.

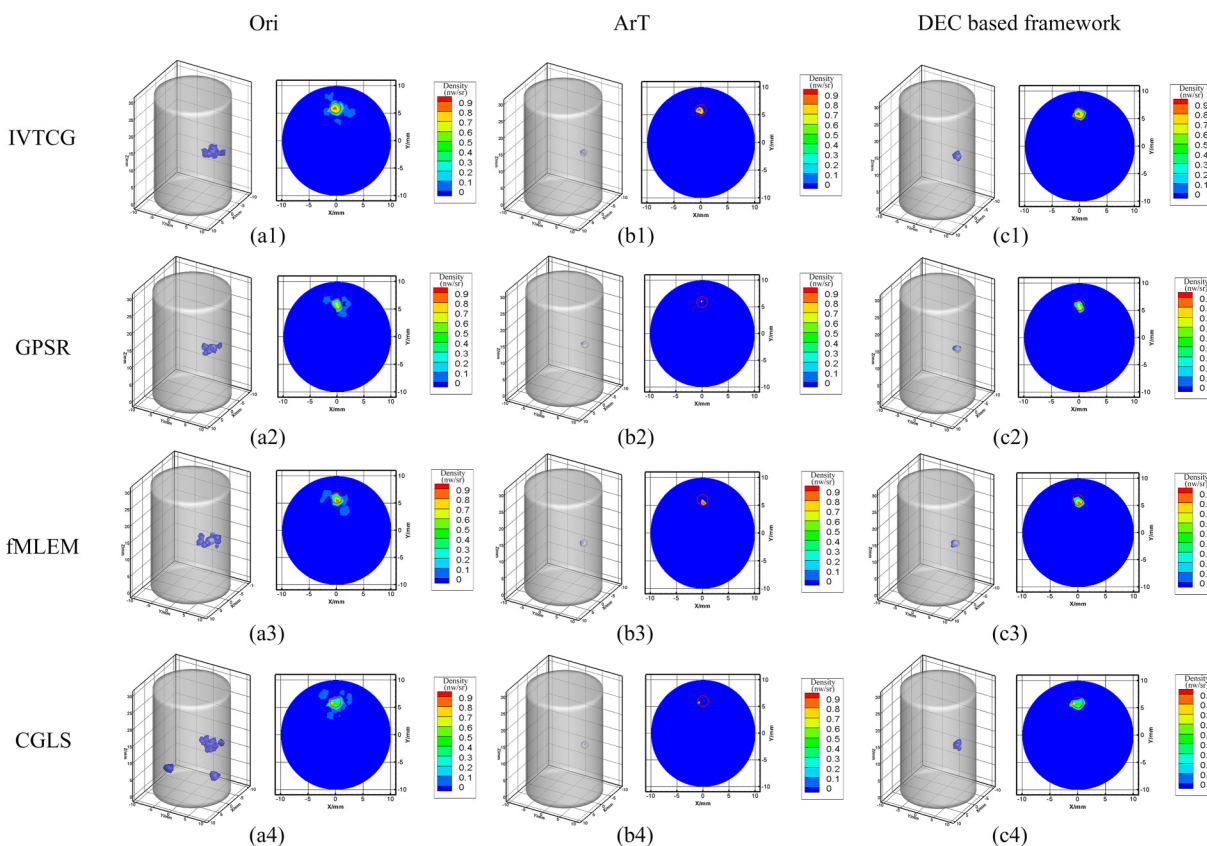


FIG. 4. Comparison results for the single-source-based experiment. (a1) presents the original result reconstructed by the IVTCG method, (b1) and (c1) are the post-processing results obtained by the ArT method and the DEC-based post-processing framework. (a2), (b2), and (c2) show the originally reconstructed result by the GPSR method, the post-processing results obtained by the ArT method, and the DEC-based post-processing framework, respectively. (a3), (b3), and (c3) are the corresponding results of the fMLEM method, and (a4), (b4), and (c4) detail the results of the CGLS method. Red circles indicate the location of the actual light source.

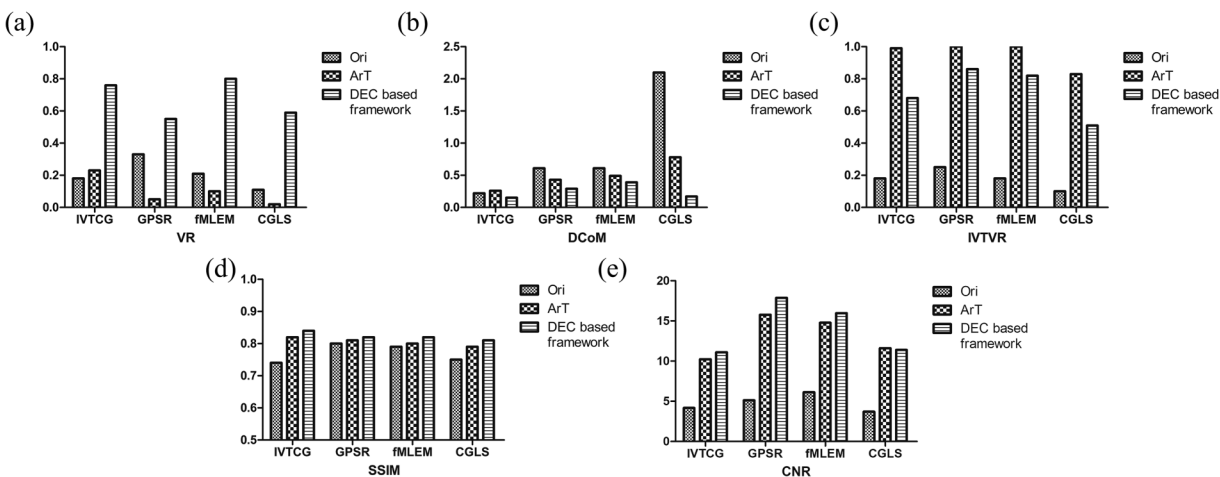


FIG. 5. Quantitative analysis of the results for the single-source-based experiment. (a)–(e) illustrate the intensity of quantitative evaluators of VR, DCoM, IVTVR, SSIM, and CNR, respectively.

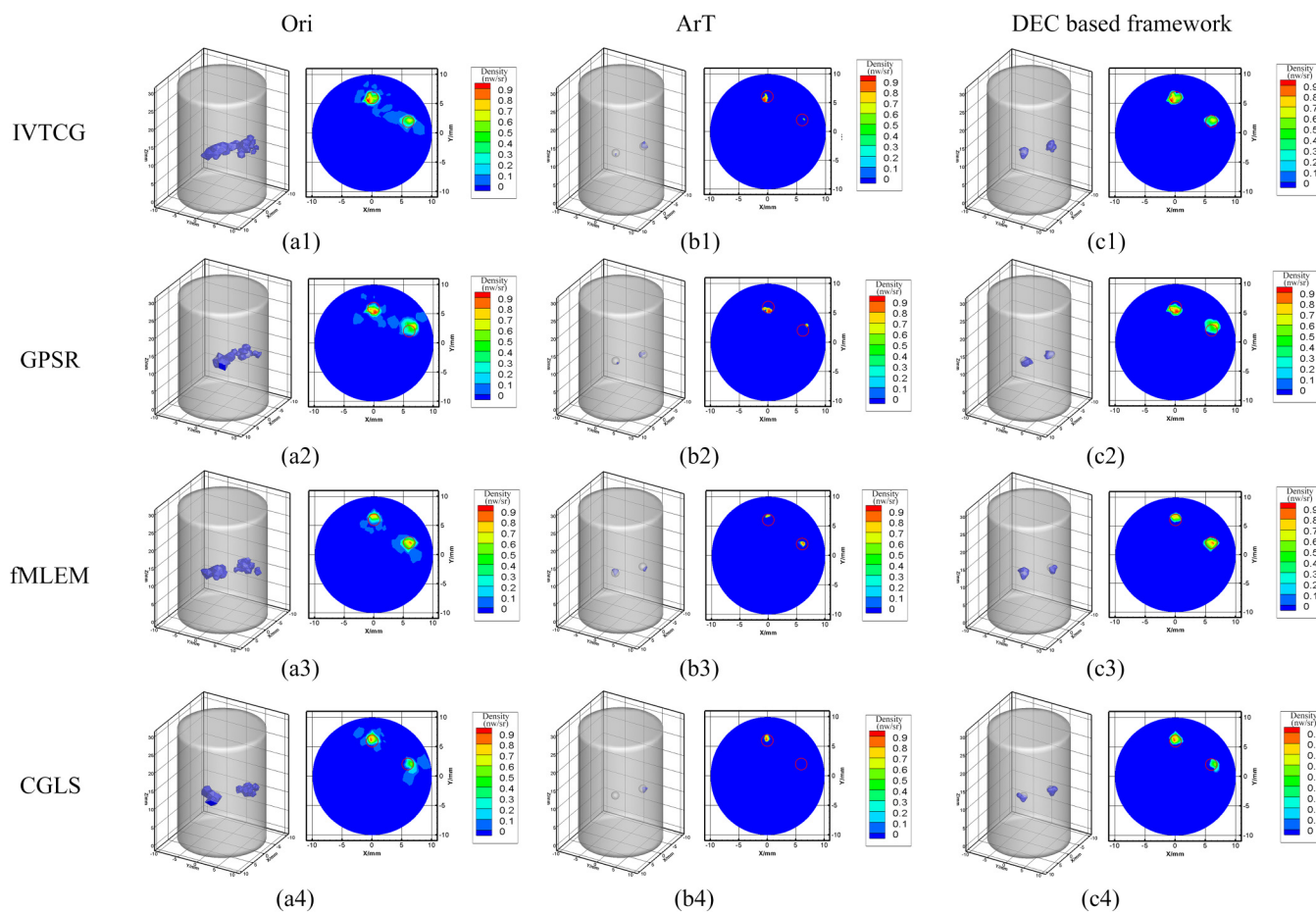


FIG. 6. Comparison results for the double-source-based experiment. (a1) presents the original result reconstructed by the IVTCG method, and (b1) and (c1) are the post-processing results obtained by the ArT method and the DEC-based post-processing framework. (a2), (b2), and (c2) show the originally reconstructed result by the GPSR method, the post-processing results obtained by the ArT method, and the DEC-based post-processing framework, respectively. (a3), (b3), and (c3) are the corresponding results of the fMLEM method, and (a4), (b4), and (c4) detail the results of the CGLS method. Red circles indicate the position of the light sources.

location of the actual light sources. Both multiple and single light sources provide the same conclusion, and there are many artifacts in the original reconstruction results. Although these artifacts were removed by the ArT method, a large amount of information in the light source region was also removed, resulting in a relatively low fidelity. Our DEC-based post-processing framework avoids this problem and fits the region of the light source very well. We further calculated the evaluator indicators defined above to analyze these results quantitatively (Fig. 7). We found that although the ArT method reduced the volume and achieved the best IVTVR intensity, the other indicators showed that the fidelity was not as good as those in our proposed framework, which made the results more similar to the light source.

C. Living mouse-based *in vivo* experiment

We further demonstrated our proposed framework's applicability and potential with an artificial-source-implanted living

animal-based *in vivo* experiment. The experimental data were collected from the CLT/micro-CT dual-modal system on an adult nude mouse, including the angle-dependent Cerenkov luminescence images and micro-CT images. After reconstruction and segmentation, the mouse's torso section with a 34 mm height was obtained [Fig. 8(a)]. The experimental mouse's main organs included the muscle, lung, heart, stomach, liver, and kidney, and the optical parameters of these organs were consistent with the literature.³² To mimic the lesion containing the radionuclide probe, we injected 11.1 MBq of ^{18}F -FDG into a thin rubber tube, and the tube was implanted into the mouse at a specific location (26, 7, 18.2 mm). Figure 8(b) is the two-dimensional cross-sectional view of Fig. 8(a) at $z = 18.2$ mm. Figure 6 shows the original reconstructed results by the IVTCG, GPSR, fMLEM, CGLS methods, and their post-processing results with the ArT and our DEC-based post-processing frameworks (Fig. 9). The quantitative analysis of these results is presented in Fig. 10, which

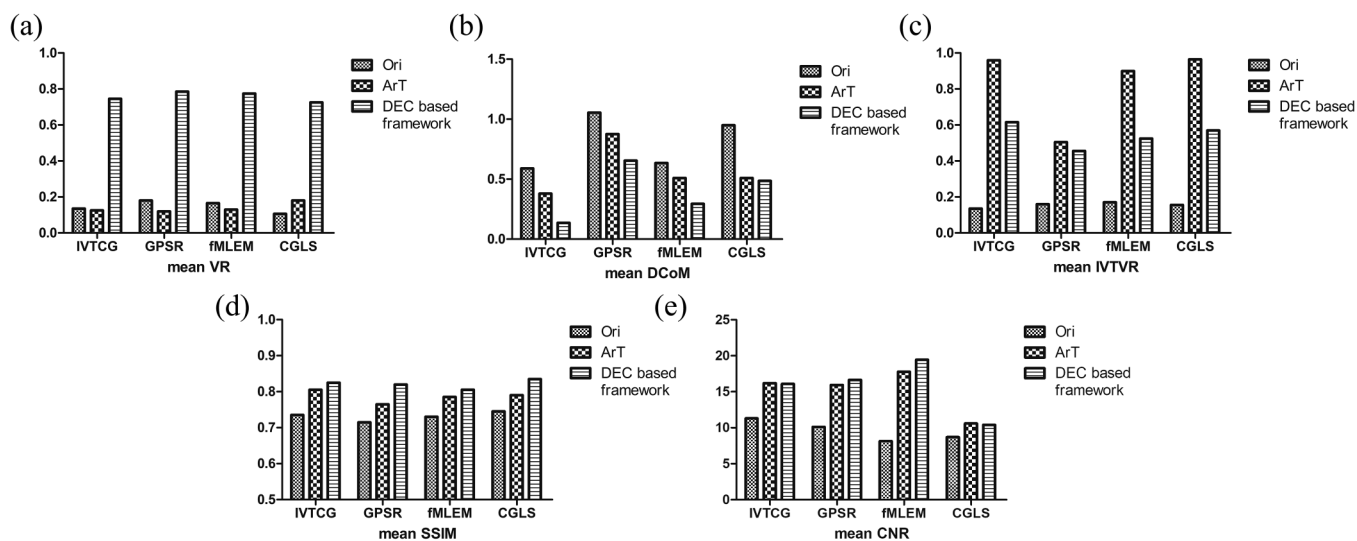


FIG. 7. Quantitative analysis of the results for the double-source-based experiment. (a)–(e) illustrate the mean intensity of quantitative evaluators of VR, DCoM, IVTVR, SSIM, and CNR, respectively.

shows that similar conclusions to the simulation experiments can be drawn. We observed that the DEC-based post-processing framework could remove the artifacts from the original reconstructed results and maintain fidelity compared with the ArT

method. Collectively, our proposed DEC-based post-processing framework has good performance in removing the artifacts and maintaining the fidelity of the reconstructed result simultaneously.

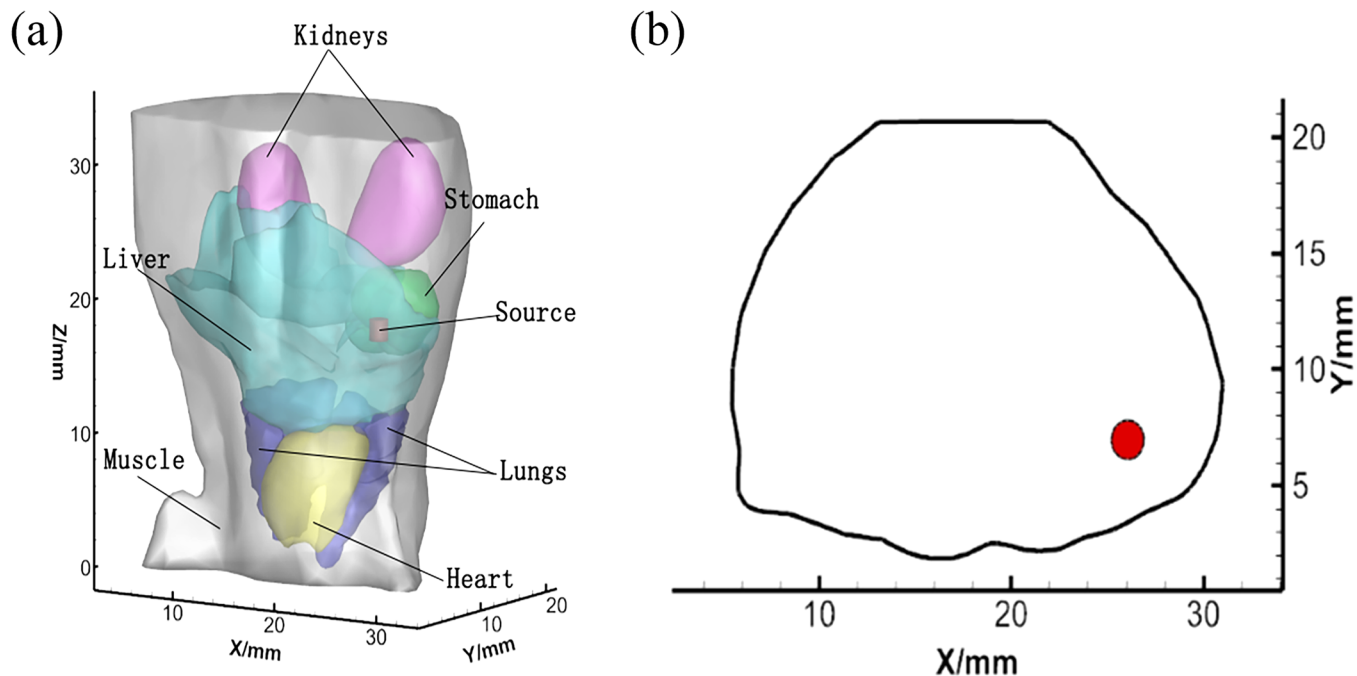


FIG. 8. Physical model used in the living mouse-based *in vivo* experiment. (a) The anatomical structure of the mouse. (b) Vertical view of the source at $z = 18.2$ mm.

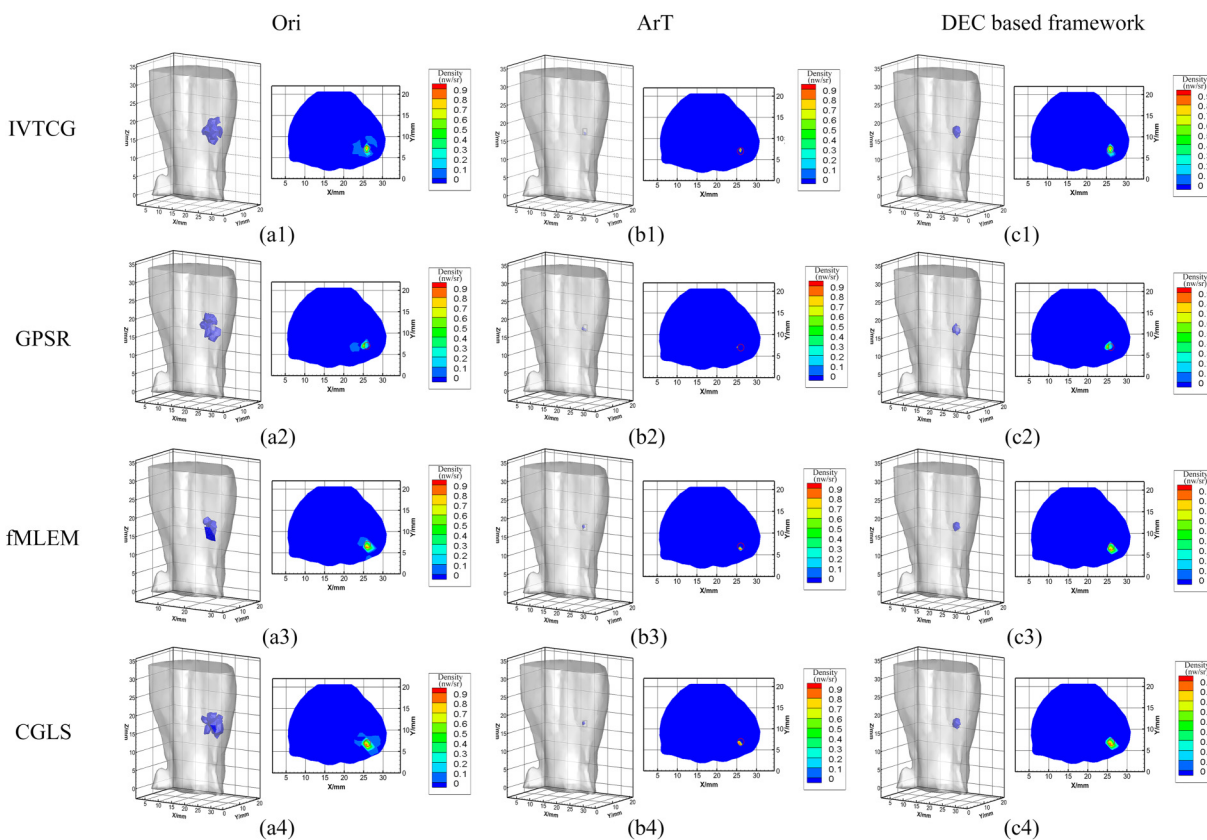


FIG. 9. Comparison results for the living mouse-based *in vivo* experiment. (a1) presents the original result reconstructed by the IVTCG method, and (b1) and (c1) are the post-processing results obtained by the ArT method and the DEC-based post-processing framework. (a2), (b2), and (c2) show the originally reconstructed result by the GPSR method, the post-processing results obtained by the ArT method, and the DEC-based post-processing framework. (a3), (b3), and (c3) are the corresponding results of the fMLEM method, and (a4), (b4), and (c4) detail the results of the CGLS method. Red circles indicate the position of the light sources.

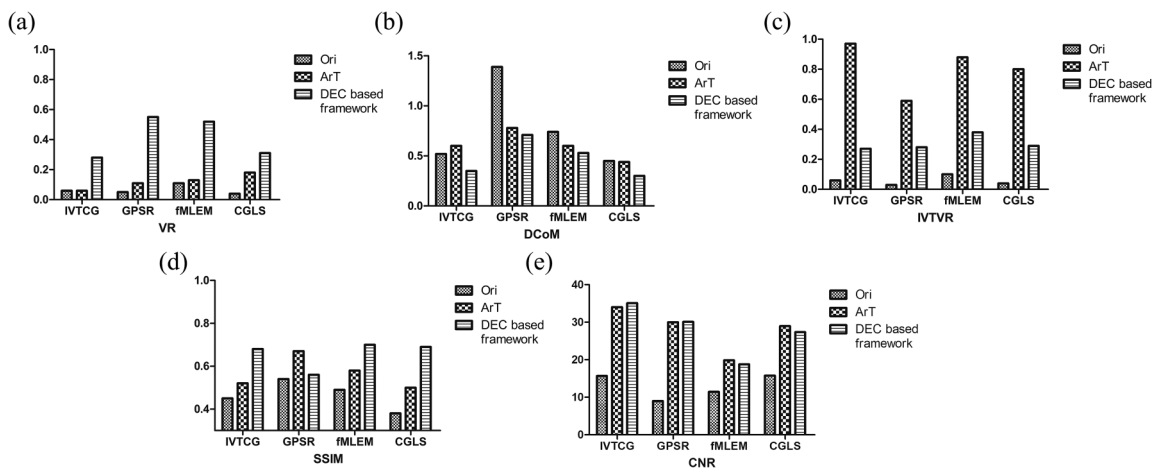


FIG. 10. Quantitative analysis of the results for living mouse-based *in vivo* experiment. (a)–(e) illustrate the intensity of quantitative evaluators of VR, DCoM, IVTVR, SSIM, and CNR, respectively.

IV. DISCUSSION AND CONCLUSION

CLT can provide a 3D distribution of radionuclide probes inside an object, and this has been used in drug development and small animal research. However, the inverse reconstruction of CLT is highly ill-posed. Low spatial resolution and the presence of light source artifacts are two direct consequences of this ill-posedness. To address these problems, a deep unsupervised clustering-based post-processing framework was reported for CLT. It can improve the spatial resolution to a certain degree and effectively eliminate these artifacts, facilitating CLT's high-fidelity reconstruction. The main advantages of our framework can be summarized as three aspects: (1) the voxelization method is used to increase the spatial resolution of the original CLT reconstructed result and to form voxelized data in the 3D space; (2) a deep unsupervised clustering algorithm is used to cluster these voxels into three groups, and the group with the highest mean intensity is chosen as the final reconstructed result; and (3) our framework is flexible, and some other more effective and efficient clustering algorithms can be incorporated in the future. The feasibility and performance of the proposed framework were demonstrated with numerical simulation and *in vivo* mouse-based experiments. The broad applicability has been validated further by post-processing of the initial reconstruction of four different reconstruction algorithms. It is hoped that the proposed framework can provide a useful tool for CLT and other optical tomographic technologies. Future work will focus on the improvement of the clustering algorithm and its application in biomedicine.

ACKNOWLEDGMENTS

This work was supported in part by the National Key R&D Program of China (Grant No. 2018YFC0910602), the National Natural Science Foundation of China (NNSFC) (No. 61701403), the China Post-doctoral Science Foundation (Grant No. 2018M643719), the Young Talent Support Program of the Shaanxi Association for Science and Technology (Grant No. 20190107), and the Scientific Research Program Funded by Shaanxi Provincial Education Department (Grant No. 18JK0767).

The authors declare that there are no conflicts of interest related to this article.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

- ¹Y. Xu, H. Liu, and Z. Cheng, *J. Nucl. Med.* **52**, 2009 (2011).
- ²Z. H. Hu, X. L. Chen, J. M. Liang, X. C. Qu, D. F. Chen, W. D. Yang, J. Wang, F. Cao, and J. Tian, *J. Appl. Phys.* **112**, 024703 (2012).

- ³X. Cao, Y. H. Zhan, X. Cao, J. M. Liang, and X. L. Chen, *Curr. Med. Imaging Rev.* **13**, 50 (2017).
- ⁴R. Robertson, M. S. Germanos, C. Li, G. S. Mitchell, S. R. Cherry, and M. D. Silva, *Phys. Med. Biol.* **54**, N355 (2009).
- ⁵Z. Hu, J. Liang, W. Yang, W. Fan, C. Li, X. Ma, X. Chen, X. Ma, X. Li, X. Qu, J. Wang, F. Cao, and J. Tian, *Opt. Express* **18**, 24441 (2010).
- ⁶A. Ruggiero, J. P. Holland, J. S. Lewis, and J. Grimm, *J. Nucl. Med.* **51**, 1123 (2010).
- ⁷X. Cao, X. Chen, F. Kang, Y. Zhan, X. Cao, J. Wang, J. Liang, and J. Tian, *ACS Appl. Mater. Interfaces* **7**, 11775 (2015).
- ⁸G. Hongbo, H. Xiaowei, L. Muhan, Z. Zeyu, H. Zhenhua, and T. Jie, *IEEE Trans. Med. Imaging* **36**, 1337 (2017).
- ⁹C. Li, G. S. Mitchell, and S. R. Cherry, *Opt. Lett.* **35**, 1109 (2010).
- ¹⁰C. Li, Y. Yang, G. S. Mitchell, and S. R. Cherry, *J. Nucl. Med.* **52**, 1268 (2011).
- ¹¹A. E. Spinelli, C. Kuo, B. W. Rice, R. Calandrino, P. Marzola, A. Sbarbati, and F. Boschi, *Opt. Express* **19**, 12605 (2011).
- ¹²J. Zhong, J. Tian, X. Yang, and C. Qin, *Ann. Biomed. Eng.* **39**, 1728 (2011).
- ¹³Z. Hu, W. Yang, X. Ma, W. Ma, X. Qu, J. Liang, J. Wang, and J. Tian, *Mol. Imaging* **12**, 173 (2013).
- ¹⁴L. Wang, X. Cao, Q. Y. Ren, X. L. Chen, and X. He, *J. Appl. Phys.* **123**, 184701 (2018).
- ¹⁵X. Cao, X. Wei, F. Yan, L. Wang, L. Su, Y. Hou, G. Geng, and X. He, *IEEE Access* **7**, 85178 (2019).
- ¹⁶X. Wei, D. Lu, X. Cao, L. Z. Su, L. Wang, H. B. Guo, Y. Q. Hou, and X. W. He, *AIP Adv.* **9**, 11 (2019).
- ¹⁷J. Zhong, C. Qin, X. Yang, S. Zhu, X. Zhang, and J. Tian, *Int. J. Biomed. Imaging* **2011**, 1 (2011).
- ¹⁸X. L. Chen, D. F. Yang, X. Cao, and J. M. Liang, *J. Nucl. Med.* **55**(Suppl. 1), 264 (2014).
- ¹⁹H. Liu, X. Yang, T. Song, C. Bao, L. Shi, Z. Hu, K. Wang, and J. Tian, *J. Biomed. Opt.* **20**, 086007 (2015).
- ²⁰H. Yi, H. Wei, J. Peng, Y. Hou, and X. He, *J. Opt. Soc. Am. A* **35**, 256 (2018).
- ²¹W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, and D. Shen, *Neuroimage* **108**, 214 (2015).
- ²²Y. LeCun, Y. Bengio, and G. Hinton, *Nature* **521**, 436 (2015).
- ²³J. Xie, R. Girshick, and A. Farhadi, *International Conference on Machine Learning* 2016, pp. 478–487.
- ²⁴D. Yang, L. Wang, D. Chen, C. Yan, X. He, J. Liang, and X. Chen, *Med. Biol. Eng. Comput.* **56**, 2067 (2018).
- ²⁵Q. Zhang, H. Zhao, D. Chen, X. Qu, X. Chen, X. W. He, W. Li, Z. Hu, J. Liu, J. Liang, and J. Tian, *Opt. Commun.* **284**, 5871 (2011).
- ²⁶M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, *IEEE J. Sel. Top. Signal Process.* **1**, 586 (2007).
- ²⁷X. He, J. Liang, X. Wang, J. Yu, X. Qu, X. Wang, Y. Hou, D. Chen, F. Liu, and J. Tian, *Opt. Express* **18**, 24825 (2010).
- ²⁸X. Chen, J. Liang, X. Cao, D. Yang, D. Chen, J. Ripoll, and J. Tian, *J. Appl. Phys.* **114**, 084701 (2013).
- ²⁹H. Yi, X. Zhang, J. Peng, F. Zhao, X. Wang, Y. Hou, D. Chen, and X. He, *Bio. Med. Res. Int.* **2016** (2016).
- ³⁰M. A. Naser and M. S. Patterson, *Biomed. Opt. Express* **2**, 3179 (2011).
- ³¹S. Ren, X. Chen, H. Wang, X. Qu, G. Wang, J. Liang, and J. Tian, *PLoS One* **8**, e61304 (2013).
- ³²H. Yi, D. Chen, W. Li, S. Zhu, X. Wang, J. Liang, and J. Tian, *J. Biomed. Opt.* **18**, 056013 (2013).