2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 979-8-3503-2447-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/EMBC40787.2023.10341089

K-CapsNet: K-Nearest Neighbor Based Convolution Capsule Network for Cerenkov Luminescence Tomography Reconstruction

Xin Cao[1], Weitong Li[1], Yi Chen[1], Mengfei Du[1], Gege Zhang[1], Jun Zhang[1], Kang Li[1,*] and Linzhi Su[1,*]

***Abstract*—Cerenkov luminescence tomography (CLT) has received significant attention as a promising imaging modality that can display the three-dimensional (3D) distribution of radioactive probes. However, the reconstruction of CLT suffers from severe ill-posed problem. It is difficult for traditional model-based method to obtain satisfactory result. Recently, deep learning-based method have shown great potential for accurate and efficient CLT reconstruction. In this study, a KNN-based convolution capsule network, named K-CapsNet, is proposed for cerenkov luminescence tomography. In K-CapsNet, the surface photon intensity is encoded in capsule form. The KNN-based convolution and K-means clustering are proposed for efficient encoding. Numerical simulation experiments have been carried out to verify the performance of K-CapsNet, and the results show that it performs superior in source localization and morphological restoration compared with existing methods.**

***Clinical Relevance*—As proved by experiments, the proposed K-CapsNet can achieve superior reconstruction accuracy, which is helpful for tumor detection, cancer drug therapy monitoring, and other applications.**

I. Introduction

Cerenkov luminescence imaging (CLI) is a promising optical molecular imaging modality, which is highly sensitive and low-cost. It has been widely used for tumor imaging, therapy monitoring, and other preclinical studies [1-3]. As an extension of CLI, cerenkov luminescence tomography (CLT) can retrieve the 3D distribution information of radioactive probes in biological tissue from the surface optical signal, which is conducive for further clinical application [4].

However, CLT is still facing some challenging issues. Due to the severe light scattering effect, CLT reconstruction is a highly ill-posed problem, and it is difficult to obtain satisfactory results. To address the issues with CLT reconstruction, some regularization methods have been introduced, including L_1-norm regularization (Lasso method), L_2-norm regularization (Tikhonov method), and Lp-norm ($0 < p < 1$) regularization [5-7]. Besides, some prior information was introduced to improve the reconstruction performance, such as the structural information from computed tomography (CT) and magnetic resonance imaging (MRI) [8, 9]. Despite the improvements obtained through these methods, the approximation error between the linear photon propagation model and the true nonlinear photon propagation process remains a significant challenge.

Recently, the neural network has been widely used in optical molecular imaging. Compared with the traditional method, the neural network directly learns the mapping relation between the surface photon intensity and the internal radioactive source. The neural network effectively avoids the deviation between the simplified photon propagation model and the realistic process of light propagation. In 2019, a multilayer fully connected neural network (MFCNN) was first applied to reconstruct the distribution of the radioactive source [10]. Also, a CLT reconstruction framework based on a stacked denoising auto-encoder was proposed, which indeed achieved a good performance in CLT reconstruction [11]. For the reconstruction in fluorescence molecular tomography and bioluminescence tomography, Meng *et al.* proposed a K-Nearest Neighbor-based locally connected network, which cascaded two sub-network for preliminary results and fine-tuning, respectively [12]. Yu *et al.* proposed a one-dimensional convolutional neural network, which had fewer parameters and higher learning efficiency [13].

In this paper, to enhance reconstruction quality, a KNN-based convolution capsule network is proposed for CLT reconstruction, named K-CapsNet. In K-CapsNet, KNN-based convolution is introduced to extract the feature of surface photon intensity. K-means clustering is employed to reduce the number of parameters. Based on the capsule network structure, the surface photon intensity information is encoded in capsule form, and the capsules will be fed into multi fully connected layers for further reconstruction. Experiments revealed that K-CapsNet achieved higher reconstruction accuracy compared with other methods.

The rest of this paper is structured as follows: In section II, the details of K-CapsNet are introduced. In section III, the numerical simulation experiments are illustrated to prove the effectiveness of our network. In section IV, the conclusion is illustrated.

II. Method

A. Model-based CLT reconstruction

For model-based CLT reconstruction, considering scattering effects are dominant over absorption effects in the detectable CLT spectrum, the diffusion approximation to the radiative transfer equation is widely used to describe the light

This work was supported in part by the National Key Research and Development Program of China (2019YFC1521102, 2019YFC1521103); Key Research and Development Program of Shaanxi Province (2019GY215, 2021ZDLSF06-04); National Natural Science Foundation of China (61701403, 61806164); China Postdoctoral Science Foundation (2018M643719); Graduate Innovation Program of Northwest University (CX2023185).

1 is with the School of Information Sciences and Technology, Northwest University, Xi'an, 710127, China

* is corresponding author (likang@nwu.edu.cn, sulinzhi029@163.com).

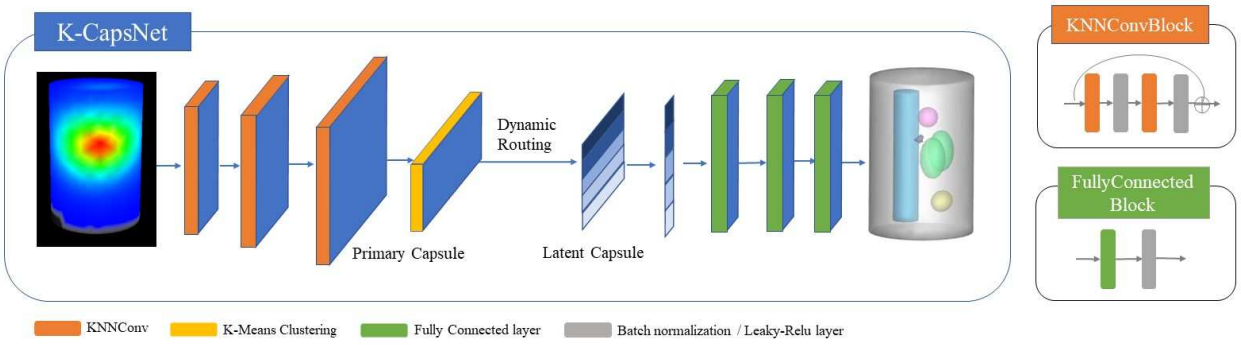

Fig.1. Network Architecture

propagation process [4]. Based on finite element analysis, a linear relation between the surface photon intensity and the distribution of the radioactive source is established, which can be defined as follows:

$$\Phi = AX \tag{1}$$

where Φ is the photon intensity of the surface. A is the system matrix determined by the characteristics of biological organs. X is the distribution of the radioactive source.

B. Deep learning-based CLT reconstruction

Deep learning-based CLT reconstruction uses neural network to establish the mapping relation between the photon intensity of surface and the radioactive source directly, which can be described as follows:

$$min\|f_{nn}(\Phi|\theta) - X\|_2^2 \tag{2}$$

where f_{nn} denotes the reconstruction neural network. θ denotes network weight which is iteratively updated. Φ denotes the surface photon intensity as an input to the network. X denotes the actual distribution of the radioactive source.

C. The architecture of K-CapsNet

K-CapsNet is based on capsule network structure, which is illustrated in Fig. 1. Firstly, three KNNConvBlocks are used to extract the feature of surface photon intensity. Each KNNConvBlock comprises two KNNConv layers, each followed by a batch normalization and a Leaky-Relu layer. KNNConv layer is designed to make use of structural information to extract features. As shown in Fig. 2 (a), the convolution domain of KNNConv is a set of adjacent nodes in three-dimensional space, allowing it to effectively extract features from neighboring nodes for each node. During implementation, the feature of each node is concatenated with the feature of its adjacent nodes, followed by a convolution, as shown in Fig. 2 (c). The adjacent relationship is established by conducting a K-nearest neighbor search to identify the nearest neighbor in 3D coordinate space, with k being set to 9.

After three KNNConvBlocks, K-means clustering is used as a pooling operation, assigning each node in the permissible region to a neighborhood. The k is set to 512. The cluster's feature is calculated as the average of the node's feature in the cluster. In K-CapsNet, the feature of each cluster serves as a primary capsule. A squash function is applied as an activation function after the clustering, which is defined as follows:

$$v = \frac{\|s\|^2}{1 + \|s\|^2} \cdot \frac{s}{\|s\|} \tag{3}$$

where s represents the input and v represents the vector output of the capsule. The primary capsules are clustered into latent capsules through the dynamic routing algorithm. Dynamic routing is an iterative algorithm, which ensures the outputs of primary capsules are sent to the proper latent capsules. The detail of the dynamic routing algorithm can be found in [14].

Then, the latent capsules are transposed and concatenated without masking before sent into three FullyConnectedBlocks. Each FullyConnectedBlock includes a fully-connected layer, a batch normalization, and a Leaky-Relu layer. The output of the FullyConnectedBlocks is the reconstructed radioactive source. The network configuration is shown in Table I. N is the number of nodes in the permissible region.

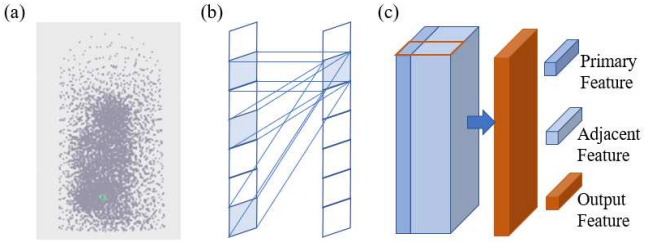

Fig. 2. The detail of KNNConv (a) the convolution domain of KNNConv in three-dimensional space (b) the convolution domain of KNNConv on input vector (c) implementation explanation of KNNConv.

TABLE I. THE NETWORK CONFIGURATION

	Input	**KNNConv**	**KNNConv**
Feature map	$N \times 1$	$N \times 8$	$N \times 32$
	KNNConv	**Primary Capsule**	**Latent Capsule**
Feature map	$N \times 128$	512×128	$32 \times 64/2048 \times 1$ (concatenated)
	Fully Connected layer	**Fully Connected layer**	**Fully Connected layer**
Feature map	2048×1	$N \times 1$	$N \times 1$

D. Implementation detail and evaluation metrics

The training and test of K-CapsNet were implemented using Pytorch and Python 3.7. All operation was performed on a personal computer with an AMD Ryzen 7 1700 Eight-Core Processor 3.00 GHz CPU and a NVIDIA GeForce GTX 1080 Ti GPU. The optimizer of K-CapsNet was Adam with a learning rate of 0.001. Mean Square Error (MSE) Loss was adopted as a loss function.

To quantify the performance of K-CapsNet, location error (LE), Dice coefficient, and signal-to-noise ratio (SNR) were used.

Location error is defined as the position error between the reconstructed radioactive source and the actual radioactive source. It can be defined as:

$$LE = \|L_r - L_0\|_2 \quad (4)$$

where L_r represents the center of the reconstructed radioactive source and L_0 represents the center of the actual radioactive source.

Dice coefficient evaluates the similarity between the reconstructed source region and the actual source region:

$$Dice = \frac{2|S_{rec} \cap S_{ac}|}{|S_{rec}| + |S_{ac}|} \quad (5)$$

where S_{rec} represents the reconstructed source region and S_{ac} represents the actual source region.

The signal-to-noise ratio (SNR) is used to measure the visual quality, which can be defined as:

$$SNR = \frac{\sum_{i=1}^{n} u_i^2}{\sum_{i=1}^{n}(u_i - u_i')^2} \quad (6)$$

where u_i' represents the energy of the i^{th} node in the reconstruction result, u^i represents the energy of the i^{th} node in the original numerical model. n is the total number of nodes in the numerical model.

III. Experiments and Results

To evaluate the performance of K-CapsNet, both single-source and dual-source numerical simulation experiments were conducted. The MFCNN method and Incomplete variable truncated conjugate gradient (IVTCG) method were taken for comparison.

In the numerical simulations, a heterogeneous cylindrical phantom was used to simulate the mouse body. The phantom consists of five kinds of organs: heart, lungs, bone, liver, and muscle, which are exhibited in Fig. 3. All the experiments are based on a single spectrum (650nm). Optical parameters of all organs are presented in Table II, which are obtained from [15].

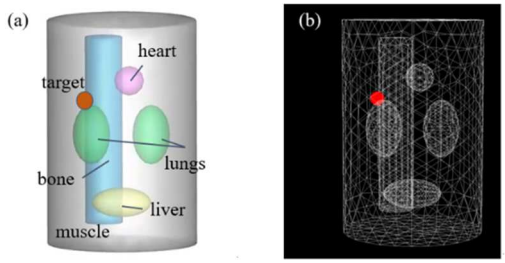

Fig. 3. The numerical phantom. (a) shows a single-source phantom, and (b) shows the standard mesh.

TABLE II. Optical parameters of different organs.

Component	Absorption coefficient $\mu_a\ (mm^{-1})$	Scattering coefficient $\mu_s\ (mm^{-1})$
Muscle	0.016	0.510
Heart	0.011	1.053
Lungs	0.036	2.246
Liver	0.012	2.472
Bone	0.021	2.864

The numerical phantom was discretized into a standard mesh with 4626 nodes. A small spherical radioactive source with 1mm diameter was set in the model to present the tumor. As a data-driven method, plenty of simulation data was collected using the Monte Carlo method. All the simulation was conducted in Molecular Optical Simulation Environment (MOSE 2.3) [16]. In this work, 221 single-source samples and have been simulated. and the corresponding data from two randomly selected single-source samples were combined to create 3000 dual-source samples. Of these samples, 80% were randomly chosen for training purposes, while the remainder were reserved for testing.

A. Single source simulation result

For the single-source experiment, the relative result was shown in Table III and Fig. 4. The source was set at (-1, -1, 15) mm. From Table III, K-CapsNet exhibited the minimum LE (0.23mm), maximum Dice (0.77) and SNR (1.56) among three methods, which indicated that our method performs better in localization and shape recovery compared with MFCNN and IVTCG. Fig. 4 shows the visualization result of K-CapsNet, MFCNN, and IVTCG respectively.

TABLE III. Quantitative results of single-source simulation.

Method	LE (mm)	Dice	SNR
K-CapsNet	0.23	0.77	1.56
MFCNN	0.58	0.67	1.39
IVTCG	1.10	0.19	0.80

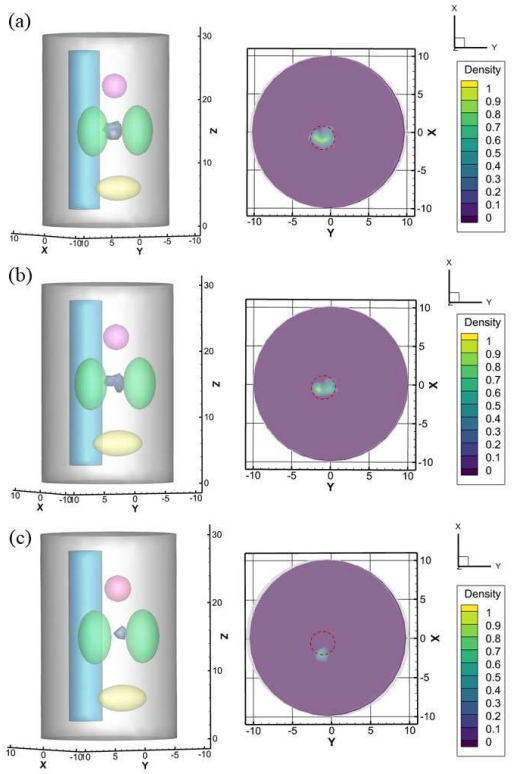

Fig. 4. Comparison results for the single-source experiments. (a) presents the result reconstructed by K-CapsNet, (b) and (c) are the results obtained by the MFCNN method and IVTCG method respectively. The actual position and size of source is indicated by red circle.

B. Dual-source simulation result

For the dual-source experiment, the relative result was shown in Table IV and Fig. 5. The centers of the dual source were set at (-5, -4, 4) mm, and (-6, 1, 8) mm respectively. The quantitative results are exhibited in Table IV, which illustrates that our network still obtained the minimum LE (0.66mm), the maximum Dice (0.63), and SNR (1.27). It can be inferred from Table IV that the reconstruction result of K-CapsNet is closer to the real source, and the performance of K-CapsNet is better than the other two methods. Fig. 5 exhibits the visualization of the dual-source reconstruction results and suggests that both

reconstructed radioactive sources are easy to distinguish with clear shapes.

TABLE IV. QUANTITATIVE COMPARISON OF DUAL-SOURCE SIMULATION.

Method	LE (mm)	Dice	SNR
K-CapsNet	0.66	0.63	1.27
MFCNN	0.89	0.53	1.10
IVTCG	1.04	0.24	1.09

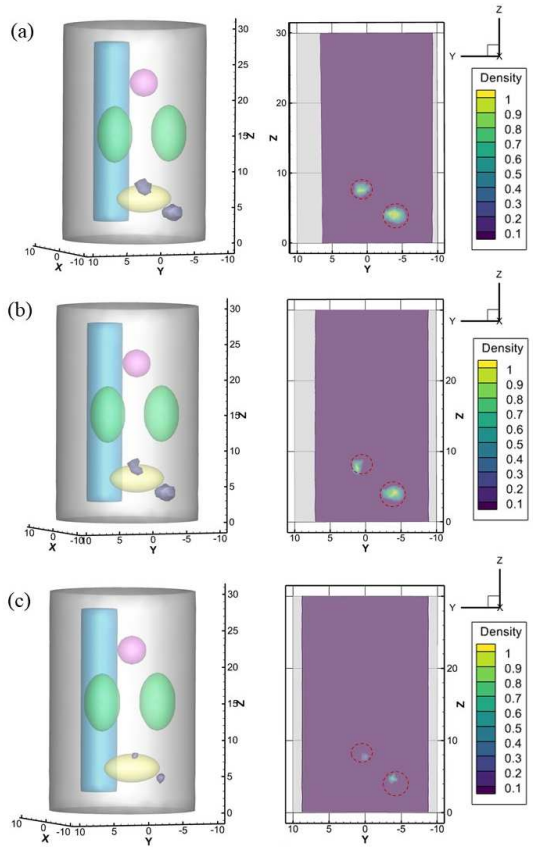

Fig. 5. Comparison results for the dual-source experiments. (a) presents the result reconstructed by K-CapsNet, (b) and (c) are the results obtained by the MFCNN method and IVTCG method respectively. The actual position and size of the source is indicated by red circle.

IV. CONCLUSION

In this study, a KNN-based convolution capsule network, named K-CapsNet, is proposed to enhance the performance of CLT reconstruction. K-CapsNet is a capsule network model, mainly consisting of KNNConvBlock and FullyConnected-Block. By using KNN-based convolution, K-CapsNet makes good use of structural information to extract features. By applying K-means clustering, K-CapsNet reduces input spaces.

To verify the performance of our network, single-source numerical experiment and dual-source numerical experiment have been done. Compared with MFCNN method and IVTCG method, K-CapsNet performs better on localization and shape recovery.

However, some issues remain unresolved in this work. For example, the KNN-based convolution is memory-consumed. The division between the standard mesh and the actual structure introduces extra error in CLT reconstruction. Besides, no *in vivo* experiments have been conducted to validate the feasibility of K-CapsNet. We will continue to address these issues in future studies.

REFERENCES

[1] R. Robertson, M. S. Germanos, C. Li, G. S. Mitchell, S. R. Cherry, and M. D. Silva, "Optical imaging of Cerenkov light generation from positron-emitting radiotracers," *Phys Med Biol,* vol. 54, no. 16, pp. N355-65, Aug 21, 2009.

[2] J. Axelsson, and J. Krohn, "Cerenkov Luminescence Imaging for Accurate Placement of Radioactive Plaques in Episcleral Brachytherapy of Intraocular Tumors," *Invest Ophthalmol Vis Sci,* vol. 56, no. 12, pp. 7362-8, Nov, 2015.

[3] D. Fan, X. Zhang, L. Zhong, X. Liu, Y. Sun, H. Zhao, B. Jia, Z. Liu, Z. Zhu, J. Shi, and F. Wang, "(68)Ga-labeled 3PRGD2 for dual PET and Cerenkov luminescence imaging of orthotopic human glioblastoma," *Bioconjug Chem,* vol. 26, no. 6, pp. 1054-60, Jun 17, 2015.

[4] C. Li, G. S. Mitchell, and S. R. Cherry, "Cerenkov luminescence tomography for small-animal imaging," *Opt Lett,* vol. 35, no. 7, pp. 1109-11, Apr 1, 2010.

[5] M. Cai, Z. Zhang, X. Shi, J. Yang, Z. Hu, and J. Tian, "Non-negative Iterative Convex Refinement Approach for Accurate and Robust Reconstruction in Cerenkov Luminescence Tomography," *IEEE Trans Med Imaging*, Apr 21, 2020.

[6] G. Hongbo, H. Xiaowei, L. Muhan, Z. Zeyu, H. Zhenhua, and T. Jie, "Weight Multispectral Reconstruction Strategy for Enhanced Reconstruction Accuracy and Stability With Cerenkov Luminescence Tomography," *IEEE Trans Med Imaging,* vol. 36, no. 6, pp. 1337-1346, Jun, 2017.

[7] H. Guo, Z. Hu, X. He, X. Zhang, M. Liu, Z. Zhang, X. Shi, S. Zheng, and J. Tian, "Non-convex sparse regularization approach framework for high multiple-source resolution in cerenkov luminescence tomography," *Optics Express,* vol. 25, no. 23, pp. 28068-28085, 2017.

[8] X. Ding, K. Wang, B. Jie, Y. Luo, Z. Hu, and J. Tian, "Probability method for Cerenkov luminescence tomography based on conformance error minimization," *Biomed Opt Express,* vol. 5, no. 7, pp. 2091-112, Jul 1, 2014.

[9] M. Liu, H. Guo, H. Liu, Z. Zhang, C. Chi, H. Hui, D. Dong, Z. Hu, and J. Tian, "In vivo pentamodal tomographic imaging for small animals," *Biomedical optics express,* vol. 8, no. 3, pp. 1356-1371, 2017.

[10] Z. Zhang, M. Cai, Y. Gao, X. Shi, X. Zhang, Z. Hu, and J. Tian, "A novel Cerenkov luminescence tomography approach using multilayer fully connected neural network," *Phys Med Biol,* vol. 64, no. 24, pp. 245010, Dec 19, 2019.

[11] X. Cao, X. Wei, F. Yan, L. Wang, L. Su, Y. Hou, G. Geng, and X. He, "A Novel Stacked Denoising Autoencoder-Based Reconstruction Framework for Cerenkov Luminescence Tomography," *IEEE Access,* vol. 7, pp. 85178-85189, 2019.

[12] H. Meng, Y. Gao, X. Yang, K. Wang, and J. Tian, "K-nearest Neighbor Based Locally Connected Network for Fast Morphological Reconstruction in Fluorescence Molecular Tomography," *IEEE Trans Med Imaging*, Apr 3, 2020.

[13] J. Yu, C. Dai, X. He, H. Guo, S. Sun, and Y. Liu, "Bioluminescence Tomography Based on One-Dimensional Convolutional Neural Networks," *Frontiers in Oncology*, pp. 4182, 2021.

[14] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," *Advances in neural information processing systems,* vol. 30, 2017.

[15] X. Cao, J. Zhang, J. Yang, C. Fan, F. Zhao, W. Zhou, L. Wang, G. Geng, M. Zhou, and X. Chen, "A deep unsupervised clustering-based post-processing framework for high-fidelity Cerenkov luminescence tomography," *Journal of Applied Physics,* vol. 128, no. 19, pp. 193104, 2020.

[16] S. Ren, X. Chen, H. Wang, X. Qu, G. Wang, J. Liang, and J. Tian, "Molecular Optical Simulation Environment (MOSE): a platform for the simulation of light propagation in turbid media," *PLoS One,* vol. 8, no. 4, pp. e61304, 2013.