2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

End-To-End Bioluminescence Tomography Reconstruction Based On Convolution Neural Network Scheme

Shuangchen Li[1,2], Xuelei He[1,2], Heng Zhang[1,2], Hongbo Guo[1,2,*], and Xiaowei He[1,2,*]

Abstract— Bioluminescence tomography (BLT) has received a lot of attention as an important technique in bio-optical imaging. Compared with traditional methods, neural network methods have the advantages of fast reconstruction speed and support for batch processing. In this paper, we propose a end-to-end BLT reconstruction based on convolution neural networks scheme. First, 3000 datasets with single source and dual sources projection were conducted by Monte Carlo method, respectively. And three convolution neural networks (VGGNet, ResNet, and DenseNet) were adopted to feature extraction. Then, the filtered features were used as input to the multi-layer perceptron (MLP) to predict the source location. The results of numerical simulation and simulation experiments show, compared with traditional methods, the advantages of our method are including high reconstruction accuracy, faster reconstruction, few parameters, simple reconstruction process and support for batch processing.

I. INTRODUCTION

As a non-invasive, highly sensitive, low-cost optical imaging modality, Bioluminescence tomography (BLT) allows to obtain 3-D distribution information and quantitatively result of the internal tumor tissue in cell and molecular lever by collecting the surface light flux distribution, and applying accurate light transfer model and inverse reconstruction method [1]. It is widely used for non-invasive visualization of tumor early detection, drug development, efficacy assessment and other preclinical studies.In theory, RTE can accurately depict photon propagation in diffusive media, but it is a complicated integro-differential equation. The diffusion equation(DE) [2] [3] [4] have been widely used as an approximation to the RTE equation for the tissues with high absorption and low scattering tissue. However, for more common used tissue, inherently contains system errors caused by DE is unavoidable for application of BLT.

Due to the non-linearity of light scattering in tissues and the effect of noise from the sounding camera, the inversion problem of BLT becomes an ill-posed problem. To deal with this problem, many reconstruction methods based on regularization have been proposed to improve the reconstruction results,such as L_2-norm,L_1/L_p-norm ($0<p<1$) and total variation norm methods [5] [6] [7] [8] [9] [10] [11]. Although these methods can deal with the ill-posedness of FMT inverse problem, however, the over-smoothness of L_2-norm results in blurred or spread targets with the loss of high-frequency feature in the reconstructed images, meanwhile L_1 and $L_p(0<p<1)$ norm may lose the edge information due to the over-sparsity in reconstruction. TV regularization-related optimizations are some of the most difficult optimizations to be solved computationally due to the non-smooth and nondifferentiable property of the TV regularizer. Besides that, in practice the results of methods based on regulazation have strong dependence on the selection of parameters. Which will influence the accuracy of 3-D distribution and quantitatively result of internal tumor.

With the rise of neural network algorithms in recent years, more and more studies have tried to use neural network algorithms to solve the inverse problem of BLT, and good results have also been achieved. For example, the fully connected network solution scheme proposed by Gao et al [12], the solution scheme based on 3-D en-decoder proposed by Guo et al [13], and the solution scheme based on recurrent neural network proposed by Huang et al [14]. The reconstruction results proved that neural network can achieve better reconstruction image contrast and localization accuracy with low computation time. Unlike traditional regularized reconstruction algorithms are based on a theory-driven algorithmic model, deep learning algorithms are based on a data-driven algorithmic model. Which can directly construct the reverse photon propagation by learning the nonlinear mapping relationship between the surface photon density and the biological source density, without using the forward photon propagation model and the reverse reconstruction method. Compared with traditional methods, neural network algorithms have the advantages of high reconstruction accuracy and fast reconstruction speed, and can effectively avoid errors caused by the inaccuracy of the transmission model.

Inspired by the research of neural network, an end-to-end bioluminescence tomography reconstruction based on convolution neural network scheme is supposed in this paper. In this scheme, first, the 6000 cases of surface BLT simulations samples were constructed by using

*This research is supported by the National Natural Science Foundation of China, Grant/Award Numbers:(61971350, 11871321, 61901374, 61906154, 82071914); Postdoctoral Innovative Talents Support Program, Grant/Award Number: (BX20180254); Project funded by China Post-doctoral Science Foundation, Grant/Award Number: (2018M643719); Science and Technology Plan Program in Xi'an of China, Grant/Award Number: (201805060ZD11CG44);

1 is with the Xi'an Key Laboratory of Radiomics and Intelligent Perception, Xi'an, China.

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

∗ is corresponding author (hexw@nwu.edu.cn),(guohb@nwu.edu.cn)

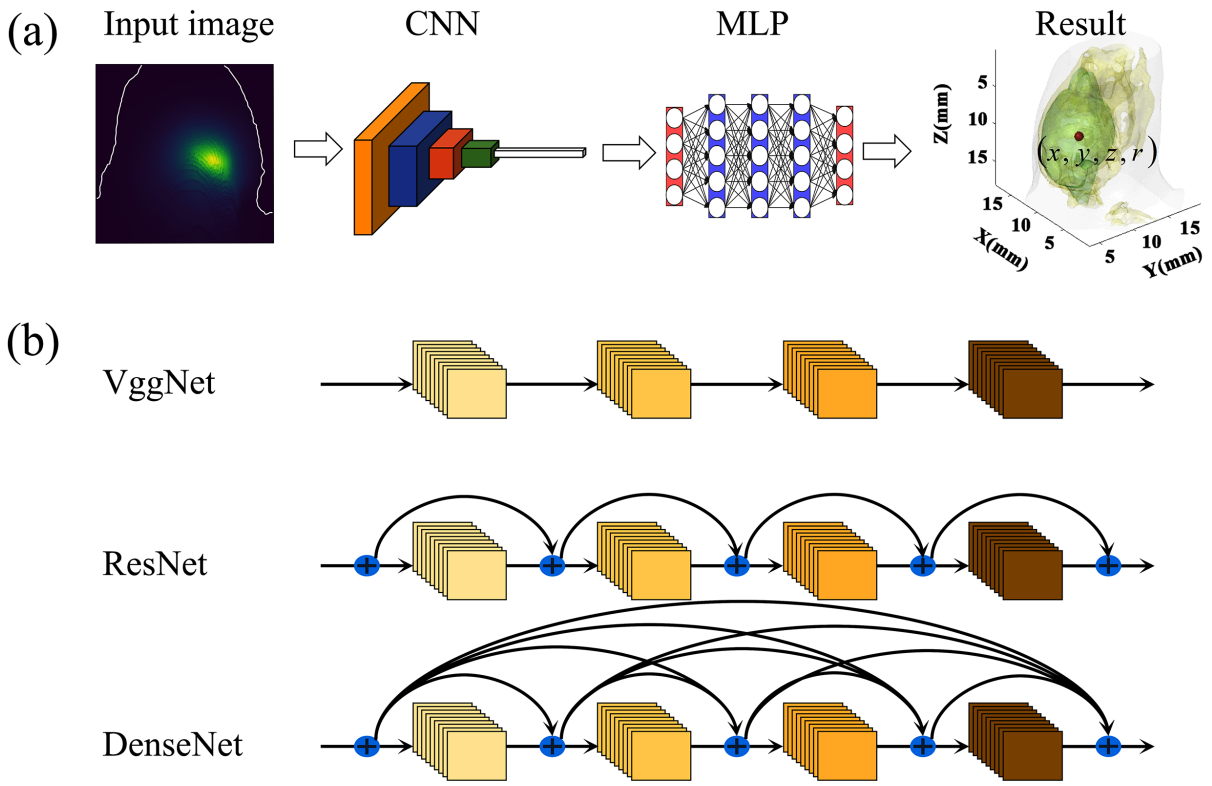

Fig. 1. Network Structure Diagram. Subgraph (a) shows basic structure diagram, from left to right are network input, convolutional neural networks(CNN), multilayer perceptron(MLP) and output result respectively. Subgraph (b) shows three different CNN structure—VggNet, ResNet and DenseNet.

Monte Carlo method. And, three convolutional neural network (VGGNet [15], ResNet [16], and DenseNet [17]) was adopted to feature extraction and screening, as the input of multilayer perceptron (MLP). Thus, it's avoid introduce numerous parameters and over come the parameter explosion problem reducing the computer memory requirement. Lastly, the proposed convolution neural network scheme were verified by a series of numerical simulations and phantom experiments. The structure of this article as follows: In the Section 2 we show the neural network model we used and the types of input and output data of the neural network; In the Section 3 we show the results of our experiments; And in the last Section we show the conclusions drawn from this experiment and some related discussions.

II. METHOD

A. Networks

Fig.1 shows the overall architecture of the network, which is consisted of convolutional neural networks(CNN) and multilayer perceptron(MLP). CNN is responsible for extracting the data features and MLP is responsible for fitting the linear mapping. For CNN part, three different networks(VggNet, ResNet and DenseNet) are compared in our study, the max-pooling operation is set after four convolutional layers for all network and the size of convolution kernel is 3*3. Especially for RetNet, the 2*2 size of convolution kernel with stride is 2 to replace the max-pooling operator and the input of each block depends on the input and output of the previous block. DenseNet is divided into four dense blocks, which is an extreme version of the residual block, every convolutional layer gets the output of all prior convolutional layers in the block as input, the max-pooling operations are divide different dense blocks. Although the three network architectures are differ in details, the basic blocks are same. There are two different operations are packaged in a single block including convolution and relu activation function. The formula for relu is shown as equation (1):

$$relu(x_i) = \begin{cases} x_i, & if \quad x_i > 0 \\ 0, & if \quad x_i < 0 \end{cases} \tag{1}$$

For MLP part, we build a four-layer perceptron model including an input layer, an output layer and two hidden layers. To ensure the normalized source vector can be fitted correctly, the sigmoid activation function is used in the hidden layers. The sigmoid function is shown as equation (2):

$$sigmoid(x_i) = \frac{1}{1 + e^{x_i}} \tag{2}$$

Meanwhile, the x_i means the output of the $i-th$ layer in the network. To ensure the stability of the inverse propagation, the *he* initialization method [18] is used to

define the weight parameters in networks. The root mean square error (NMSE) is used as the loss function. The *he* initialization method and NMSE formula are shown as equation (3) and equation (4), where N_i means the parameter quantity in the $i-th$ layer and the W_i^j(j = output or label) means output or label in the network.

$$W \sim N\left(0, \sqrt{\frac{2}{N_{\mathrm{i}}}}\right) \tag{3}$$

$$loss = \frac{1}{N} \cdot \sum_{i=1}^{n} \left(w_i^{output} - w_i^{label}\right)^2 \tag{4}$$

B. Datasets

Since the surface photon density and the source information is a 3-D field distribution function in space, it is difficult to use as inputs or labels for the neural network directly, so we can convert it to a unique mapping in low-dimensional space and describe the source information as a simple vector equivalently. To validate our proposed method, we divide the mouse head into 300 slices and remove 120 useless slices without the brain tissue at the top in practice. The surface photon density on both sides are smaller than front in the mouse head and the structure of mouse head is a continuous non-concave surface, the positively oriented projection of the surface photon density satisfies the unique mapping condition. Considering a spherical source was used, we choose to uniquely describe the source by a simple set of vectors including the position and radius of source. Obviously, the relationship between this vectors and source also satisfies the unique mapping. We used Monte Carlo method (MCX) [19] to simulate the diffusion of spherical sources with different sizes in a mouse brain with size 18*20.8*18 mm. And the voxels data were divided into voxels with a volume resolution of 0.1 mm. Besides, we obtained 3000 sets of single-source datas and 3000 sets of dual-source datas. And the network input is a projection image, which is the surface photon density were projected into the 2-D plane at a fixed angle.

C. Training

The optimizer is Adam with learning rate is 1e-5 and the epoch number is 300. We took 10 percents data from the total data as validation set and the rest of the data for training randomly to ensure the validity of the conclusions drawn from our experiments.

III. EXPERIMENT

To evaluate the performance of different networks on the reconstruction effect. We compared three different convolution neural networks, VGGNet, ResNet, and DenseNet with FISTA both in single and dual sources simulation. Based on the vector which consist of two sets of source coordinates and radius. For single source, the first four values and the last four values of the vector are the same value. We use the location distance (LE) and root mean square error (NMSE) as evaluation metrics, where the formulas for LE is shown in Equation (5). It is worth noting that the form of NMSE in the evaluation indicators and the form of the loss function of network are same.

$$LE = \sqrt{\left(x_{pre} - x_{true}\right)^2 + \left(y_{pre} - y_{true}\right)^2 + \left(z_{pre} - z_{true}\right)^2} \tag{5}$$

A. Single source experiment

For the single source experiment, the experiment results can be shown as Fig.2. In the Fig.2, the visualization reconstruction results with different algorithms are given, including the information between the real source (red) and the reconstructed source (blue) and more details at higher resolutions are shown in mouse brain space respectively. The Table.I shows the single source information, evaluation indicators and different algorithms. The results show our method has higher accuracy than FISTA algorithm. Due to the high sparsity of the reconstruction results, the FISTA algorithm cannot reconstruct a spherical source with 0.2 mm radius. For the neural network, the accuracy basically increases with the increase of the source radius, for large sources, the reconstruction accuracy of DenseNet is particularly outstanding, while the source radius change has the least effect on the reconstruction accuracy for ResNet, but has a greater effect on VggNet and DenseNet. Due to use the convolution instead of max-pooling to down sampling, which increases the stability of the model in ResNet. Due to more residual connections are used in DenseNet, which to make it can fit more deeper feature. So, the reconstruction accuracy of DenseNet is better than ResNet and VggNet.

B. Double source experiment

For the dual source experiment, the experiment results can be shown as Fig.3. In the Fig.3, the visualization reconstruction results with different algorithms are given, including the information between the real source (red) and the reconstructed source (blue) and more details at higher resolutions are shown in mouse brain space respectively. The Table.II shows the dual source information, evaluation indicators and different algorithms. The results show our method has higher resolution than FISTA algorithm. Same as in the single source experiment, due to the effect of source sparsity, the reconstruction accuracy of FISTA algorithm for large source is higher than small source. FISTA cannot reconstruct small sources with a radius of 0.2 mm. In neural network, the effect of VggNet becomes worse with the source radius increase because of the interweaving part of the two sources increases with the the source radius. Larger the interweaving area is, the more blurred the information of a single source is. Due to introduce the residual links, more deep features in the intertwined region are extracted, which makes DenseNet and ResNet perform better than VggNet in this regard obviously.

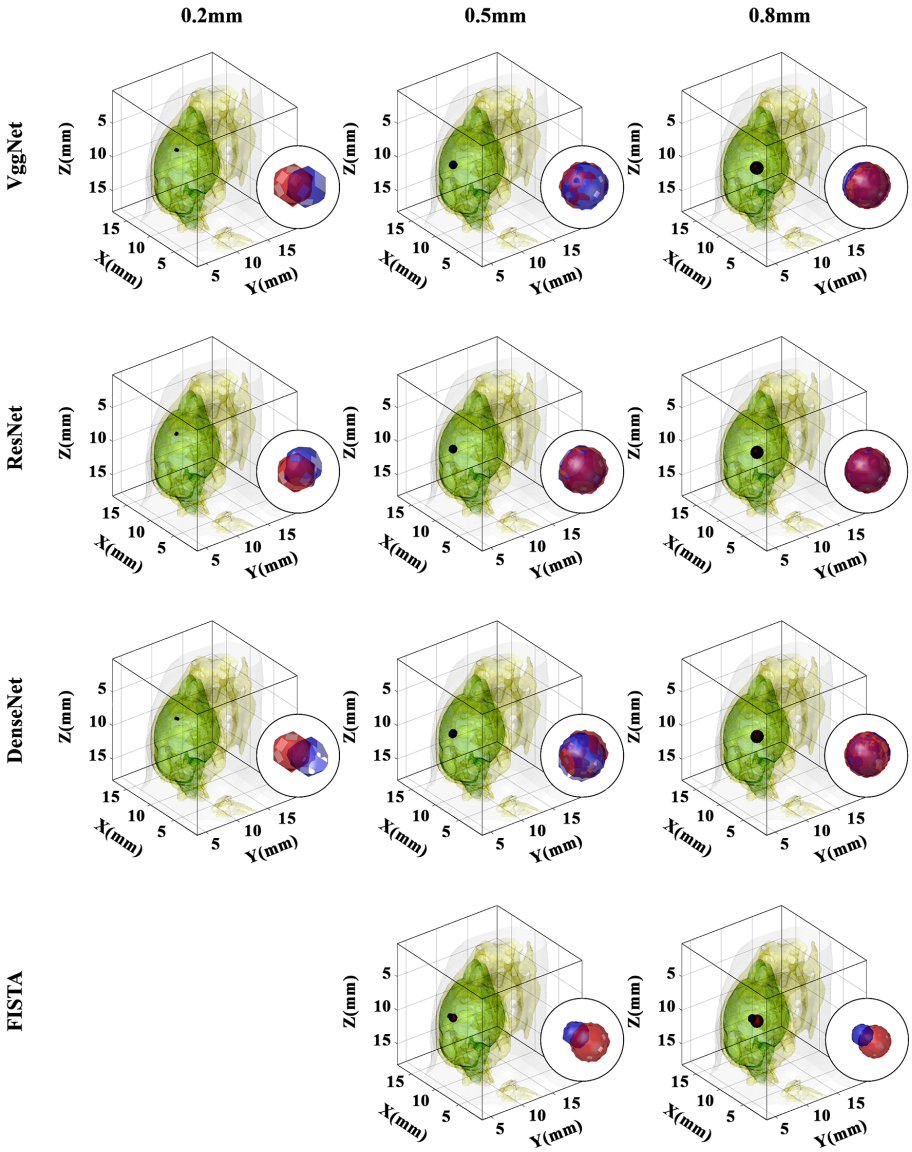

Fig. 2. Reconstruction results of a single source. Which red is the real source and blue is the algorithm to reconstruct the source. From top to bottom are the reconstruction results of VGG, ResNet, DenseNet and FISTA algorithms. The radusi of the test source data used from left to right are 0.2mm, 0.5mm and 0.8mm respectively.

It can be concluded that the method in this paper outperforms the traditional algorithm in results.

IV. CONCLUSION

In this paper, we propose an end-to-end bioluminescence tomography reconstruction based on convolution neural network scheme by building three different convolutional neural networks. Single-source and double-sources experiments were conducted to verify the feasibility and effectiveness by comparing with the conventional FISTA algorithm. The main results indicate that our proposed framenwork have a great improvement in positioning accuracy and volume recovery.

Our framework return the coordinates of the source by convolutional operation on the projection image of the light intensity distribution on the object surface directly. It release the problem of ill-poseness in BLT reconstruction and avoids the problems of slow iteration and complex parameter adjustment in traditional methods effectively. Compared to the finite element framwork, our proposed framework based on the voxel-based MCX method can overcome the limitation of the tetrahedral mesh.Comparative experiments with the FISTA algorithm, which confirm that the reconstruction results of the finite element-based mesh scheme are constrained by the mesh density (the FISTA algorithm cannot reconstruct a source with a radius of 0.2 mm).And for the comparison of different networks, we find that ResNet and DenseNet have good robustness, while the

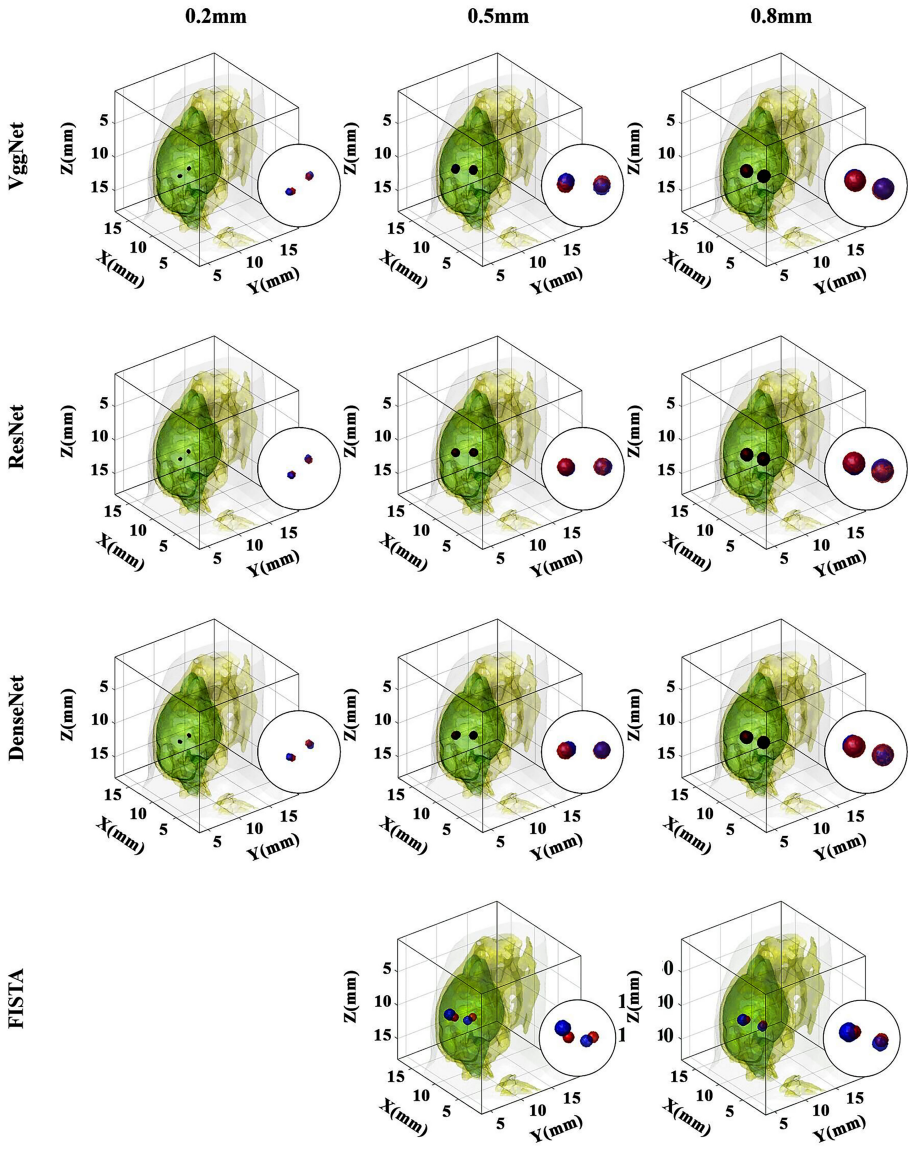

Fig. 3. Reconstruction results of dual sources. Which red is the real source and blue of the algorithm to reconstruct the source. From top to bottom are the reconstruction results of VGG, ResNet, DenseNet and FISTA algorithms. The radius of the test source data used from left to right are 0.2mm, 0.5mm and 0.8mm respectively.

results of DenseNet are closest to the real situation. DenseNet has denser residual edge connections than ResNet and can extract deeper features, but these deeper features may also be more susceptible to noise. Since VGGNet has no residual edge connectivity and loses features due to the increase in the number of layers, VGGNet is much less effective than ResNet and DenseNet. On the other hand, since ResNet uses a convolution operation instead of a downsampling operation, it has more network parameters compared to VGGNet and ResNet, and therefore is not as fast to compute as VGGNet and DenseNet. From a series experiments, the results show that the error of our proposed method is between 1 and 2 voxels, and the variance is within an acceptable range. Meanwhile, the steps of grid dissection are avoided, simplifying the operation process.

However, in our research, output only reflect the position and size of the spherical sources, which has some limitation for describing the irregular or non-spherical source. In futhure's works, our frameworks will be improved in output form, generalizability and interpretability. Besides that,some in vivo experiments will be conducted to test the sensitivity of our proposed method.

TABLE I

The differences between the reconstruction methods based on three different neural networks and the conventional FISTA algorithm under different signle source corresponding to the three metrics (NMSE and LE).

	Radius(mm)	NMSE(mm)	LE(mm)
VGGNet	0.2	0.22284	0.22284
	0.5	0.02867	0.02648
	0.8	0.11739	0.11732
ResNet	0.2	0.22635	0.22634
	0.5	0.13416	0.13415
	0.8	0.13014	0.12664
DenseNet	0.2	0.2369	0.2369
	0.5	0.08925	0.08815
	0.8	0.05545	0.05172
FISTA	0.2	——	——
	0.5	0.52453	0.49621
	0.8	1.19066	1.158

TABLE II

The differences between the reconstruction methods based on three different neural networks and the conventional FISTA algorithm under different double source corresponding to the three metrics (NMSE and LE).

	Radius (mm)	NMSE(mm) source1	LE(mm) source1	NMSE(mm) source2	LE(mm) source2
VGGNet	0.2	0.12801	0.12801	0.17529	0.17529
	0.5	0.16971	0.16971	0.169	0.16899
	0.8	0.24162	0.24145	0.3942	0.39407
ResNet	0.2	0.11	0.11	0.1117	0.1117
	0.5	0.09144	0.09144	0.18446	0.18446
	0.8	0.13067	0.13012	0.09719	0.09657
DenseNet	0.2	0.19021	0.19021	0.19605	0.19604
	0.5	0.18358	0.182	0.22466	0.22327
	0.8	0.10579	0.09037	0.26139	0.25554
FISTA	0.2	——	——	——	——
	0.5	1.39981	1.39005	0.60433	0.59649
	0.8	0.30192	0.27154	0.45573	0.41095

References

[1] R. Weissleder and M. J. Pittet, "Imaging in the era of molecular oncology." Nature, vol. 452, no. 7187, pp. 580–589, 2008.

[2] R. Gong, B. Hofmann, and Y. Zhang, "A new class of accelerated regularization methods, with application to bioluminescence tomography," 2019.

[3] L. Yin, K. Wang, T. Tong, Y. An, H. Meng, X. Yang, and J. Tian, "Improved block sparse bayesian learning method using k-nearest neighbor strategy for accurate tumor morphology reconstruction in bioluminescence tomography," IEEE Transactions on Biomedical Engineering, vol. 67, no. 7, pp. 2023–2032, 2019.

[4] J. Feng, K. Jia, Z. Li, B. W. Pogue, M. Yang, and Y. Wang, "Bayesian sparse-based reconstruction in bioluminescence tomography improves localization accuracy and reduces computational time," Journal of biophotonics, vol. 11, no. 4, p. e201700214, 2018.

[5] J. Ye, C. Chi, Z. Xue, P. Wu, Y. An, H. Xu, S. Zhang, and J. Tian, "Fast and robust reconstruction for fluorescence molecular tomography via a sparsity adaptive subspace pursuit method," Biomedical optics express, vol. 5, no. 2, pp. 387–406, 2014.

[6] X. He, J. Yu, X. Wang, H. Yi, Y. Chen, X. Song, and X. He, "Half thresholding pursuit algorithm for fluorescence molecular tomography," IEEE Transactions on Biomedical Engineering, vol. 66, no. 5, pp. 1468–1476, 2018.

[7] X. He, H. Meng, X. He, K. Wang, X. Song, and J. Tian, "Nonconvex laplacian manifold joint method for morphological reconstruction of fluorescence molecular tomography," Molecular Imaging and Biology, pp. 1–13, 2021.

[8] J. Dutta, S. Ahn, C. Li, S. R. Cherry, and R. M. Leahy, "Joint l1 and total variation regularization for fluorescence molecular tomography," Physics in Medicine & Biology, vol. 57, no. 6, p. 1459, 2012.

[9] H. Guo, L. Gao, J. Yu, X. He, H. Wang, J. Zheng, and X. Yang, "Inside cover: Sparse-graph manifold learning method for bioluminescence tomography (j. biophotonics 4/2020)," Journal of Biophotonics, vol. 13, no. 4, p. e202070011, 2020.

[10] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," IEEE Signal Processing Letters, vol. 14, no. 10, pp. 707–710, 2007.

[11] Y. Liu, H. Guo, H. Zhao, Y. Hou, and X. He, "A finite element mesh regrouping-based hybrid light transport model for enhancing the efficiency and accuracy in bioluminescence tomography," in Medical Imaging 2020: Physics of Medical Imaging, vol. 11312. International Society for Optics and Photonics, 2020, p. 113122H.

[12] Y. Gao, K. Wang, Y. An, S. Jiang, H. Meng, and J. Tian, "Nonmodel-based bioluminescence tomography using a machine-learning reconstruction strategy," Optica, vol. 5, no. 11, pp. 1451–1454, 2018.

[13] L. Guo, F. Liu, C. Cai, J. Liu, and G. Zhang, "3d deep encoder–decoder network for fluorescence molecular tomography," Optics letters, vol. 44, no. 8, pp. 1892–1895, 2019.

[14] C. Huang, H. Meng, Y. Gao, S. Jiang, K. Wang, and J. Tian, "Fast and robust reconstruction method for fluorescence molecular tomography based on deep neural network," in Imaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues XVII, vol. 10881. International Society for Optics and Photonics, 2019, p. 108811K.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[17] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," pp. 4700–4708, 2017.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," pp. 1026–1034, 2015.

[19] Y. Yuan, L. Yu, and Q. Fang, "Denoising in monte carlo photon transport simulations using gpu-accelerated adaptive nonlocal mean filter," in Optical Tomography and Spectroscopy. Optical Society of America, 2018, pp. JTh3A–41.