Spatiotemporal Attention Constrained Deep Learning Framework for Dual-Tracer PET Imaging

Dankun Lian[1], Yue Li[1], and Huafeng Liu[1,2,3(✉)]

[1] State Key Laboratory of Modern Optical Instrumentation, College of Optical Science and Engineering, Zhejiang University, Hangzhou 310027, China
liuhf@zju.edu.cn
[2] Jiaxing Key Laboratory of Photonic Sensing and Intelligent Imaging, Jiaxing 314000, China
[3] Intelligent Optics and Photonics Research Center, Jiaxing Research Institute, Zhejiang University, Jiaxing 314000, China

Abstract. Dual-tracer positron emission tomography (PET) imaging can provide the concentration distribution of two tracers in the body in a single scan, helping to better diagnose and understand diseases. However dual-tracer PET imaging separation is a challenging problem because of indistinguishable gamma photon pairs. In this work, we propose a two-dimensional convolutional network to separate the reconstructed mixed activity images, with the aid of channel attention modules to pay attention to both spatial and temporal information, which play an important role in the separation. Simulation experiments with different tracer pairs, scanning times, and phantoms are conducted to verify the generalization and robustness of the method to noise and individual differences. And its performance is also evaluated with real datasets. These results demonstrate the proposed method might have strong potential for the dual-tracer PET imaging.

Keywords: Dual-tracer PET imaging · Separation · Deep learning · Spatiotemporal information

1 Introduction

Positron Emission Tomography (PET), a powerful medical imaging technique, is often used to identify distribution of radiolabeled probes at molecular level and becomes an extremely effective diagnostic aid. Dynamic dual-tracer PET can provide more comprehensive spatiotemporal information than single-tracer PET in one scan, which saves time and cost while improving the accuracy of diagnosis and helping doctors choose more effective treatment options [1]. However, since the gamma photon pairs emitted by different tracers have the same energy (both 511 keV), it is difficult to distinguish different tracer signals in dual-tracer

G. Yang et al. (Eds.): MIUA 2022, LNCS 13413, pp. 87–100, 2022.
https://doi.org/10.1007/978-3-031-12053-4_7

PET imaging. Currently, methods for imaging the concentration distribution of two tracers in one scan are mainly divided into traditional methods and deep learning methods. Traditional methods [2–4] require additional information and are affected by tracer pairs and injection intervals. In contrast, deep learning methods do not suffer from these limitations and can achieve good separation effect because of their powerful feature learning capabilities.

At present, deep learning methods to achieve dual-tracer signal separation are mainly divided into two categories. The first method starts from mixed images, reconstructed from EM or MAP methods. Time activity curves (TACs) show the changes in the concentration value of a pixel on the image at different time frames, and TACs are extracted from these images for separation, such as stacked auto-encoder (SAE) [5], deep belief network (DBN) [6,7] and mask-based bidirectional gated recurrent unit (MB-BGRU) [8]. The limitation of these methods is to separate the signal of each pixel without considering the spatial information. Another type of method is to use a three-dimensional convolutional network to reconstruct and separate the activity images of a single tracer directly from the mixed sinogram, such as FBP-CNN [9] and multi-task learning three-dimensional convolutional network [10]. Although these methods can focus on both temporal and spatial information, the large amount of parameters in the network makes training difficult. Considering that reconstruction in a neural network consumes a lot of memory, or there are no constraints, the generalization is limited and a large amount of data is required [11]. This work proposes a method starting from the reconstructed activity images, that can focus on both temporal and spatial information, and has fewer network parameters, low training cost, and stronger generalization and robustness.

This work proposes a two-dimensional convolutional neural network based on U-net to separate two single-tracer activity images from the reconstructed dual-tracer activity images. Convolutional networks can handle spatial information. Considering that temporal information is as important as spatial information during dynamic dual-tracer PET imaging, we input the activity images of a series of time frames into the network and take the number of time frames as the initial number of feature channels, and focus on the time information by adding channel attention mechanisms to the network. Compared with 3D-CNN (the separation network part in FBP-CNN) which simultaneously focuses on spatiotemporal information, this network has ten times fewer parameters. We set up four sets of simulation experiments to verify the generalization of the network to sampling protocol, tracer pair and phantom shape, robustness to noise and kinetic parameter variation range, and the performance of the method is demonstrated with real experiments. We compare this method with MB-BGRU method and 3D-CNN method, using different evaluation metrics, our method can outperform with about 10 times faster.

2 Methodology

2.1 Dual-Tracer PET Imaging Model

The model for simultaneous injection of dual-tracer PET imaging can be described as:

$$X^{dual}(t) = GY^{dual}(t) + e(t) \tag{1}$$

$$Y^{dual}(t) = Y^{I}(t) + Y^{II}(t) \tag{2}$$

where t represents time, X^{dual} is the dual-tracer sinogram obtained by PET scan, G is the system matrix, Y^{dual} is the dual-tracer activity images, which is composed of Y^{I} and Y^{II} of the two tracers injected separately, and e is the noise generated by sampling process. This work starts from the mixed activity images reconstructed from the sinogram and uses the proposed network to obtain Y^{I} and Y^{II}.

2.2 Network Architecture

We designed the network structure on the basis of U-net [12], as shown in Fig. 1. Considering dynamic PET imaging, we input the dual-tracer activity images of the same slice at different time frames into the network. Because it is a two-dimensional convolution network, the number of feature map channels at the beginning is the number of time frames. The convolutional layers in the network are followed by a batch normalization layer and a Leaky ReLU activation layer, except the activation function of the last convolutional layer is softplus. There are two convolutional layers before and after each change in the number of feature channels. The convolutional layer will eliminate some details of the image, so we add a skip connection pointing from the input of the first convolutional layer to the output of the second convolutional layer. At the end of U-net, two channel attention modules [13] are added. Because the channel dimension is initially represented by the number of time frames, the channel attention module is designed to improve the utilization of temporal information. Finally, the network outputs a stacked activity images of the two tracers.

2.3 Loss Function and Evaluation Metrics

Loss Function. The loss function we used in the training process was composed of mean square error (MSE) and structural similarity index (SSIM), as follows:

$$L = \alpha[MSE(\hat{y}_1, y_1) + MSE(\hat{y}_2, y_2)] - \beta[ln\frac{1 + SSIM(\hat{y}_1, y_1)}{2} + ln\frac{1 + SSIM(\hat{y}_2, y_2)}{2}] \tag{3}$$

where α and β are weighting factors that balance MSE and SSIM, we tested the performance of different combinations using grid search and finally set to 1 and 0.02 respectively in the training process; $\hat{y}_1$ and $\hat{y}_2$ represent the predicted image and y_1 and y_2 are the ground truth image, where subscripts 1 and 2 represent two tracers, respectively. MSE evaluates the difference between each pixel in the predicted image and the ground truth image, SSIM measures image similarity in terms of luminance, contrast and structure. MSE and SSIM can be calculated as:

$$MSE(\hat{y}, y) = \frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{N} \tag{4}$$

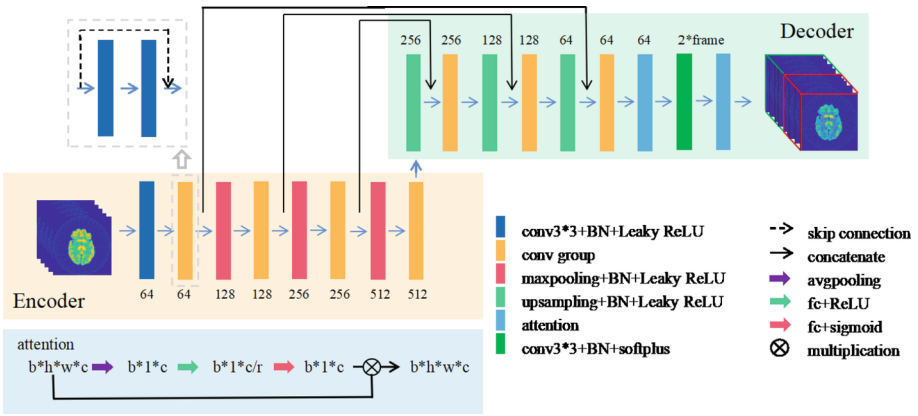


Fig. 1. A schematic diagram of the network structure, the feature channels will change each time downsampling or upsampling, and the number of feature channels in each layer is marked in the figure. Two channel attention mechanisms are added at the end of the network (see the lower left corner for details), and finally a stacked activity images of the two tracers is output.

$$SSIM(\hat{y}, y) = \frac{(2\mu_y\mu_{\hat{y}} + c_1)(2\sigma_{y\hat{y}} + c_2)}{(\mu_y^2 + \mu_{\hat{y}}^2 + c_1)(\sigma_y^2\sigma_{\hat{y}}^2 + c_2)} \tag{5}$$

where N is the number of pixels, μ_y, σ_y and $\sigma_{y\hat{y}}$ are mean of y, standard deviation of y and covariance of y and $\hat{y}$, c_1 and c_2 are constants used to prevent the denominator from being zero, related to the range of pixel value in ground truth image.

Evaluation Metrics. We used MSE, Multi-scale SSIM (MS-SSIM) [14] and Peak Signal to Noise Ratio (PSNR) to assess the separated activity images quality. MS-SSIM considers the resolution, and is a combination of SSIM calculated separately after scaling the two pictures from large to small. PSNR is defined as:

$$PSNR = 10 \cdot log_{10}[\frac{y_{max}^2}{MSE(\hat{y}, y)}] \tag{6}$$

where y_{max} is the maximum value of the image.

The image has singular points due to noise, so before calculating the loss function and evaluating the results, the image should be normalized to a common range. The normalization method is as follows:

$$y_{norm} = \frac{y - min(y)}{max(y) - min(y)} \tag{7}$$

where min(·) gets the minimum value of the image pixel value, max(·) gets the maximum value of the image pixel value.

3 Experiments and Results

3.1 Simulation Datasets and Implementation Details

The three-dimensional Zubal phantom (128 pixel × 128 pixel × 40 slice) [15] is mainly used in the simulation, and this phantom has five regions of interest (ROI). We used the parallel compartment model to simulate the dynamic spatial distribution of the tracer. The kinetic parameters describing the velocity of the tracer movement between different ROIs are obtained in real experiments in the literature [16–18]. To generate more data, we randomly select kinetic parameters from a Gaussian distribution with mean values of the true experimental values. In the simulation experiment, we used different tracer pairs with different half-lives. Tracers with long half-lives require longer scanning time. Each scanning time has 18 time frames. The details of sampling protocol setting are shown in Table 1. Each set of experiments generates 1200 sets of dynamic PET data, each set of data contains 18 time frames, of which 960 sets are used for training, 120 sets are used for validation, and 120 sets are used for testing. The training set, validation set and test set have different kinetic parameter. Since the sinogram needs a system matrix to reconstruct the activity images, which consumes a lot of memory in the network, we use the Filtered Back-projection (FBP) reconstruction algorithm to obtain the activity image (128 pixel × 128 pixel × 18 frame) and input it into the network for subsequent separation.

Table 1. Details of sampling protocol setting

Tracer	Scanning time	Scanning protocol
^{18}F-FDG/^{11}C-FMZ	40 min	$2 \times 60\,s + 2 \times 90\,s + 14 \times 150\,s$
	50 min	$3 \times 60\,s + 7 \times 140\,s + 8 \times 230\,s$
	60 min	$2 \times 60\,s + 6 \times 180\,s + 10 \times 240\,s$
^{11}C-FMZ/^{11}C-acetate	30 min	$4 \times 30\,s + 12 \times 110\,s + 2 \times 180\,s$

We input the dual-tracer activity images into the network for separation, and the respective activity images of the two tracers are used as the ground truth of the network. The optimizer of the network is Adam, the learning rate is set to 0.0002 during the training process, the batch size is 8, and a total of 100 epochs are trained.

3.2 Simulation Experiments

^{18}F-FDG+^{11}C-FMZ. For the first and second sets of experiments, we selected the tracer pair ^{18}F-FDG and ^{11}C-FMZ. Due to the large difference in half-life of these two tracers (110 min and 20.4 min), we set the scanning time to 40 min, 50 min and 60 min, and generated 1200 sets of data (10 group kinetic parameters

× 3 scanning time × 40 slice) to form the datasets with a ratio of 8:1:1 for training, validation and testing. We trained and tested it using proposed network and 3D-CNN network, considering that this phantom has the first two ROIs per slice, we calculated the average pixel value on the first two ROIs of each frame for both methods, resulting in six TAC curves. Figure 2 shows these TAC curves, it can be seen that both methods apply different sampling protocols, but for the first tracer (^{18}F-FDG), the 3D-CNN method deviates seriously, our method is closer to the ground truth, and it can be seen that using the channel attention module can better focus on temporal information.

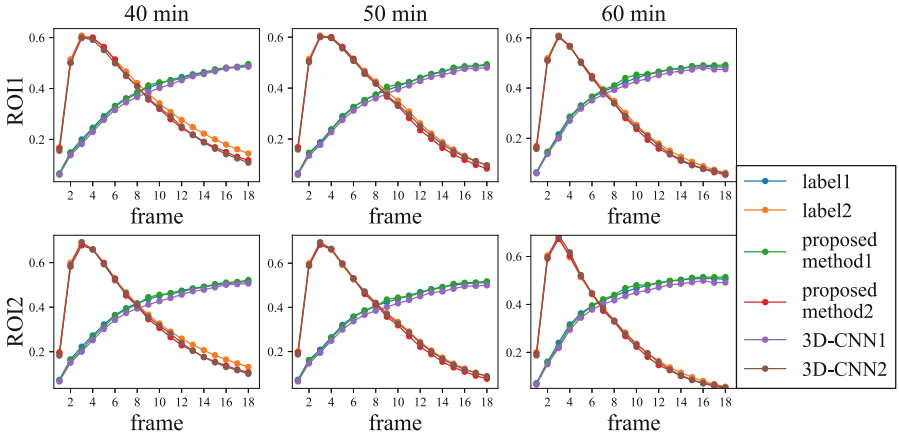
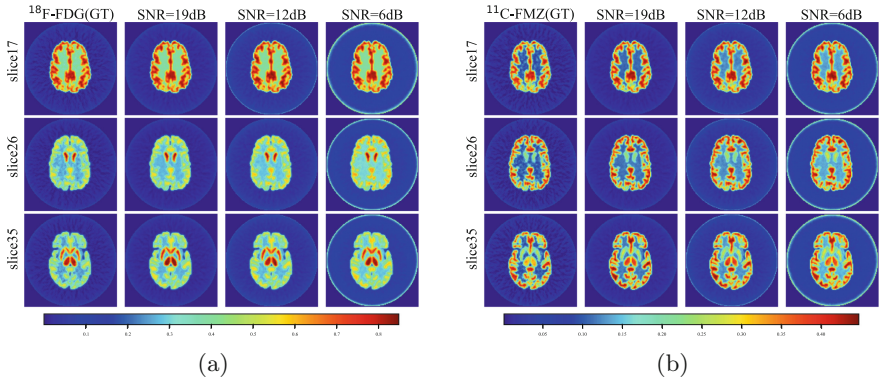


Fig. 2. TAC curve comparison chart. Each row is an ROI area, followed by 40 min, 50 min and 60 min. There are six TAC curves in each figure, blue and orange represent ground truth, green and red represent the results of the proposed method, and purple and brown represent 3D-CNN results. (Color figure online)

We superimpose random noise on the simulated sinogram, the signal-to-noise ratios (SNR) are 19 dB, 12 dB and 6 dB respectively, and trained and tested them respectively. Each data set has 1200 sets of data as before. The results are shown in Fig. 3, which shows the activity images of the 17th, 26th and 35th slices, with two ROIs, three ROIs, and five ROIs, respectively. It can be seen that as the signal-to-noise ratio decreases, the noise artifacts become larger, but still present a clear and smooth image, which is not much different from the ground truth. It reflects the high robustness of the network to noise.

^{11}C-FMZ+^{11}C-acetate. For the third set of experiments, we selected the tracer pair ^{11}C-FMZ and ^{11}C-acetate. The difference from the previous group is that the half-life of the previous group of tracers is about 5 times different, this group of tracers has the same half-life. Because of the short half-life of this set of tracer pair, we set the scanning time to 30 min. We randomly generate varying kinetic parameters by Gaussian to simulate physiological differences between individuals. In order to explore the performance of the network in the



(a) (b)

Fig. 3. Separation results of the 17th, 26th and 35th slices at SNR of 19 dB, 12 dB and 6 dB. (a) ^{18}F-FDG, (b) ^{11}C-FMZ.

case of large individual differences, we set the variance of Gaussian random to 10%, 15% and 20% of the mean, and each group has 1200 sets of data (30 group kinetic parameters × 40 slice). Figure 4(a)–(c) shows the profile of the separation results using the proposed network and MB-BGRU network. It can be seen that when the variation range of the kinetic parameters becomes larger, the deviation between the activity images and the ground truth of the two networks becomes larger, but the proposed network has less deviation and clearer boundaries than the MB-BGRU network, which shows that the network has good robustness to the differences between individuals, demonstrating the importance of paying attention to spatial information. Figure 4(d) and (e) show the evaluation metrics MS-SSIM and PSNR of the two networks in each group of experiments. The larger the value, the better the effect. It can be seen that our method outperforms MB-BGRU in both metrics, further verifying the superiority of proposed method.

Three Two-Dimensional Phantoms. In the fourth set of experiments, we selected another three two-dimensional phantoms. They are two-dimensional Zubal complex brain phantom, Zubal thorax phantom and Hoffman simple brain phantom (64 pixel × 64 pixel). Each phantom has different ROIs, corresponding to different tissues or organs. We chose three tracer pairs ^{18}F-FDG and ^{18}F-FLT, ^{62}Cu-ATSM and ^{11}C-DTBZ, ^{18}F-FLT and ^{11}C-FMZ. These tracer pairs have the same or very different half-lives, the shorter half-life of the radionuclide adopts the shorter scanning time, and a total of 90 sets of data (10 group kinetic parameters × 3 scanning time × 3 phantoms) are generated. The ratio for training and testing is 4:1. The results are shown in Fig. 5. It can be seen that for each phantom, the network achieves a certain separation effect, which proves the generalization of the network to different organ structures.

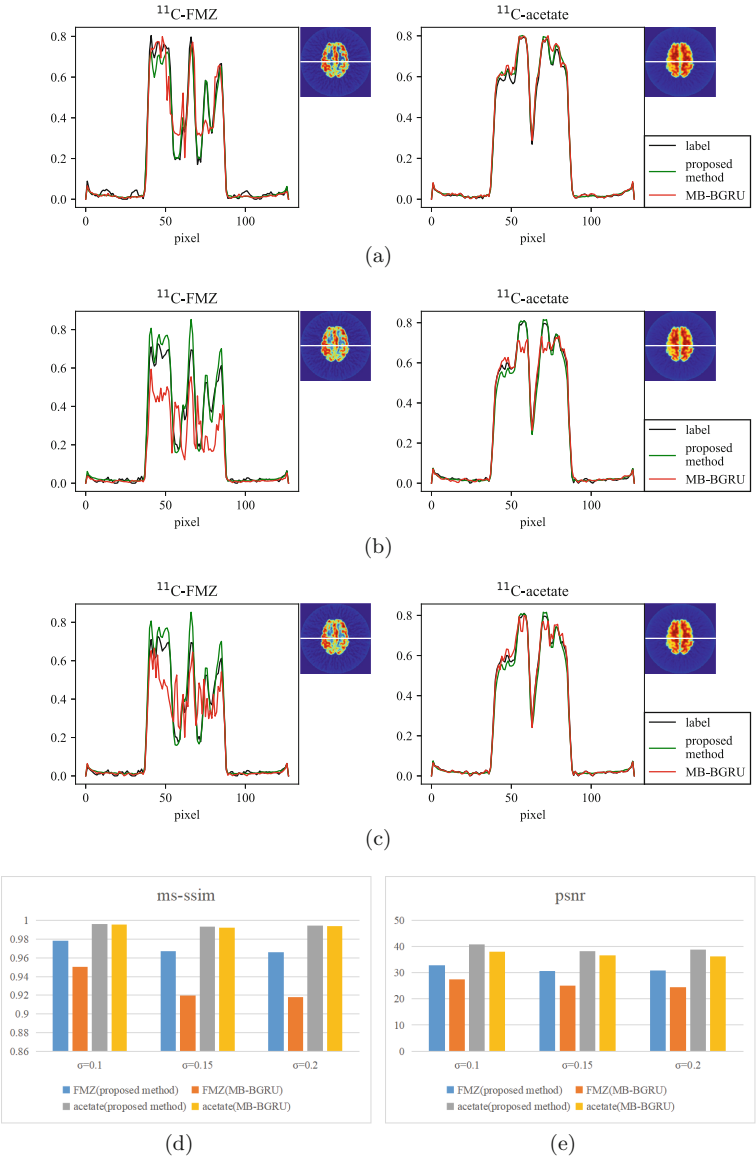


Fig. 4. Profiles of separation results using the proposed network and MB-BGRU network at different ranges of kinetic parameters. The black line represents the label, the green line represents the proposed method, and the red line represents the MB-BGRU method. The range from top to bottom is (a) 10%, (b) 15%, (c) 20%. (d) The MS-SSIM values of the proposed method and the MB-BGRU method on different tracers for different ranges of kinetic parameters. (e) Similar to (d), the value of PSNR. (Color figure online)

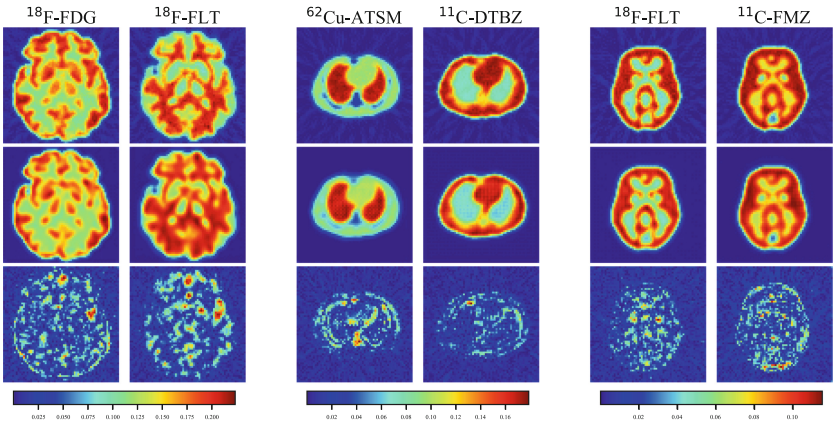


Fig. 5. The separation results of the three phantoms, the first row is the ground truth, the second row is the separation result of the proposed network, and the third row is the absolute error map.

3.3 Real Experiments

Studies have shown that in certain brain regions, mitochondria-related energy exhaustion may precede glycolysis-related hypometabolism due to pathologically confirmed early neurodegeneration in Alzheimer's disease [19]. ^{18}F-BCPP-EF can image the activity of mitochondrial complex I, and ^{18}F-FDG can measure the local brain glucose metabolism rate, so the simultaneous use of these two tracers can obtain more information in the study of neurodegenerative diseases, which has clinical significance. We used a high-resolution small animal PET scanner (SHR-38000; Hamamatsu Photonics KK, Hamamatsu, Japan) to perform dynamic PET scans of five male rhesus monkeys (rhesus macaques) weighing 4.7–8.7 kg, sequentially injected with two tracers and at intervals of more than one week to ensure complete metabolism of the tracer in the body. We used a scanning time of 120 min ($6 \times 10\,\text{s} + 2 \times 30\,\text{s} + 8 \times 60\,\text{s} + 10 \times 300\,\text{s} + 6 \times 600\,\text{s}$) to obtain 32 frames of dynamic PET data. The acquired raw data is listmode, converted to sinogram and reconstructed with Iterative 3D Dynamic Raw-Action Maximum Likelihood Algorithm (3D-DRAMA). Considering that the reconstructed image may be noisy, we performed Gaussian smoothing on the reconstruction result as the ground truth for network separation. To reduce errors caused by multiple scans, we summed the unsmoothed data from two single-tracer scans as a dual-tracer activity images. We chose the data of four monkeys as the training set and the data of the other monkey as the test set.

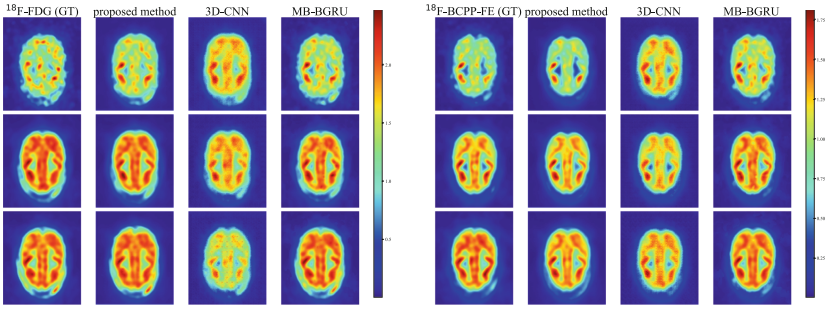


Fig. 6. Separation results of the proposed method, 3D-CNN and MB-BGRU on real data, from top to bottom are the 6th, 18th and 30th frames of the same slice. (a) ^{18}F-FDG, (b) ^{18}F-BCPP-EF.

The experimental results are shown in Fig. 6. The proposed network, 3D-CNN and MB-BGRU are used for training and testing respectively, and the activity images of frames 6, 18 and 30 are shown in turn. It can be seen that the separation results of 3D-CNN are somewhat blurred, and even obvious noise appears. Our method and MB-BGRU perform better, but for the 6th frame, the proposed method is closer to the ground truth than MB-BGRU. Table 2 shows the quantitative results of the three networks, although the proposed network has a lower average of some metrics in ^{18}F-FDG than the MB-BGRU network, it has a smaller standard deviation. At the same time, for the same amount of data, the training time is less, the speed is about 10 times that of the other two. The proposed method can obtain better quality separation results in less time.

4 Discussion

The network proposed in this work is designed on the basis of U-net, mainly adding channel attention mechanisms to focus on time information. In order to prove that the proposed method is more effective than U-net, we use the original U-net and our method to train and test on the first set of simulation data, respectively, and obtain their respective evaluation metrics as shown in Fig. 7(a)–(c). The smaller the MSE, the better, and the larger the MS-SSIM and PSNR, the better. It can be seen that the proposed method is superior to the original U-net in evaluation metrics, and has a smaller standard deviation. Figure 7(d) shows the TAC curves of ROI1 and ROI2 at different sampling times, for the tracer ^{18}F-FDG, both methods perform well, but for another tracer, our method is closer to the ground truth than the original U-net, especially after the fourth frame. The proposed method performs better on separation and also demonstrates the importance of considering temporal information.

Table 2. The evaluation metrics of the three methods in the real experiment.The value in front is the average, the ± is followed by the standard deviation

Metrics	Tracer	Proposed method	3D-CNN	MB-BGRU
MSE	^{18}F-FDG	0.0016 ± 0.00031	0.00026 ± 0.00084	0.0016 ± 0.00051
	^{18}F-BCPP-EF	$0.00085 \pm 6.56\text{e-}5$	0.0019 ± 0.00054	0.0025 ± 0.00067
MS-SSIM	^{18}F-FDG	0.952 ± 0.0024	0.923 ± 0.0067	0.956 ± 0.0061
	^{18}F-BCPP-EF	0.968 ± 0.0012	0.947 ± 0.0045	0.945 ± 0.0080
PSNR	^{18}F-FDG	28.04 ± 0.81	26.08 ± 1.23	28.24 ± 1.47
	^{18}F-BCPP-EF	30.71 ± 0.33	27.34 ± 1.20	26.20 ± 1.21

The loss function of this method uses two weighting factors α and β to balance MSE and SSIM. We set α to 1, and β to 0.1, 0.05, 0.02 and 0.01 in turn, and use loss functions of different proportions to train and test on the first set of simulation data. Figure 7(e) and (f) represent the MSE and PSNR under different proportions, it can be seen that the best performance is obtained when α is set to 1 and β is set to 0.02.

However, there are some problems to be solved. First, the method is to separate the dual-tracer signals from the reconstructed mixed activity images, so the result is limited by the accuracy of the traditional reconstruction algorithm. Second, considering the problems of tracer dose matching and physiological differences caused by multiple scans, the mixed activity images that we input into the network in both simulation and real experiments are reconstructed from the addition of two tracer sinograms. However, in practice, the scanning results obtained after the simultaneous injection of two tracers may not be the simple addition of two single tracers after injection. It is possible that the concentration of one tracer is higher than the concentration of another tracer due to the dose of tracer, human absorption, etc. Third, this work is mainly to prove the availability of the proposed method in dual-tracer PET imaging. The experiments are mainly based on phantom images, and no specific clinical application is analyzed. In future work, we can consider adding reconstruction modules and constraints to the neural network to improve the accuracy of the reconstructed activity images with less memory loss, and also need to consider the injection dose ratio of the tracer concentration. At the same time, the analysis of the specific physiological structure should be considered, whether it meets the clinical requirements needs further research.

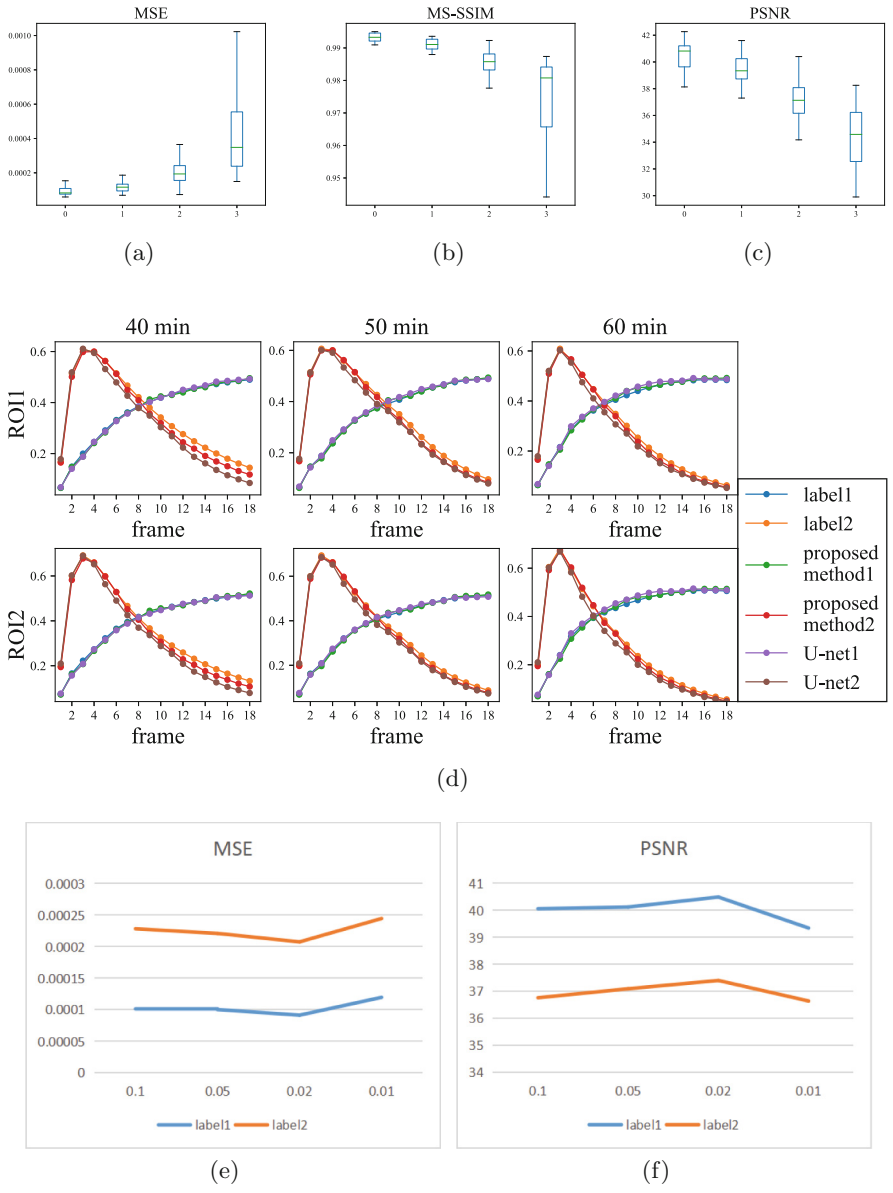


Fig. 7. Boxplots of evaluation metrics obtained using the proposed method and the original U-net, each boxplot from left to right is ^{18}F-FDG (the proposed method), ^{18}F-FDG (original U-net), ^{11}C-FMZ (proposed method), ^{11}C-FMZ (original U-net). (a) MSE, (b) MS-SSIM, (c) PSNR. (d) TAC curves of ROI1 and ROI2 under different sampling protocols, blue and orange represent ground truth, green and red represent the results of the proposed method, and purple and brown represent original U-net results. (e) and (f) represent the MSE and PSNR obtained using different scaled loss functions, and the two lines represent different tracers.

5 Conclusion

This paper proposed a deep learning network framework based on U-net to separate the mixed dual-tracer activity images into two single-tracer activity images. Compared with the method that separates TACs after extracting TACs from the mixed activity images, this method takes spatial information into account. Compared with the 3D convolutional network, by adding channel attention modules to pay attention to temporal information, it can take less time to obtain a good-quality single-tracer activity images. We verify the superiority of this method in simulation experiments and real experiments. Four sets of simulation experiments verify the generalization of the network to sampling protocols, tracer pairs, and phantom shapes, and robustness to noise and kinetic parameter variations. Real experiments also prove that this method outperforms other networks in evaluation metrics and image quality.

Acknowledgements. This work was supported by the Talent Program of Zhejiang Province (2021R51004).

References

1. Kadrmas, D.J., Hoffman, J.M.: Methodology for quantitative rapid multi-tracer pet tumor characterizations. Theranostics **3**(10), 757 (2013)
2. Huang, S., Carson, R., Hoffman, E., Kuhl, D., Phelps, M.: An investigation of a double-tracer technique for positron computerized tomography. J. Nucl. Med. **23**(9), 816–822 (1982)
3. Koeppe, R.A., Raffel, D.M., Snyder, S.E., Ficaro, E.P., Kilbourn, M.R., Kuhl, D.E.: Dual-[11c] tracer single-acquisition positron emission tomography studies. J. Cereb. Blood Flow Metab. **21**(12), 1480–1492 (2001)
4. Kudomi, N., Hayashi, T., Teramoto, N., Watabe, H., Kawachi, N., Ohta, Y., Kim, K.M., Iida, H.: Rapid quantitative measurement of CMRO2 and CBF by dual administration of 15o-labeled oxygen and water during a single pet scan - a validation study and error analysis in anesthetized monkeys. J. Cereb. Blood Flow Metab. **25**(9), 1209–1224 (2005)
5. Ruan, D., Liu, H.: Separation of a mixture of simultaneous dual-tracer pet signals: a data-driven approach. IEEE Trans. Nucl. Sci. **64**(9), 2588–2597 (2017)
6. Xu, J., Liu, H.: Deep-learning-based separation of a mixture of dual-tracer single-acquisition pet signals with equal half-lives: a simulation study. IEEE Trans. Radiat. Plasma Med. Sci. **3**(6), 649–659 (2019)
7. Qing, M., Wan, Y., Huang, W., Xu, Y., Liu, H.: Separation of dual-tracer pet signals using a deep stacking network. Nucl. Instrum. Methods Phys. Res. Sect. A **1013**, 165681 (2021)
8. Tong, J., Wang, C., Liu, H.: Temporal information-guided dynamic dual-tracer pet signal separation network. Med. Phys. (2022)
9. Xu, J., Liu, H.: Three-dimensional convolutional neural networks for simultaneous dual-tracer pet imaging. Phys. Med. Biol. **64**(18), 185016 (2019)
10. Zeng, F., Liu, H.: Dual-tracer pet image direct reconstruction and separation based on three-dimensional encoder-decoder network. In: Optics in Health Care and Biomedical Optics X, vol. 11553, p. 115530X. International Society for Optics and Photonics (2020)

11. Wang, B., Liu, H.: FBP-net for direct reconstruction of dynamic pet images. Phys. Med. Biol. **65**(23), 235008 (2020)
12. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
13. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)
14. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multiscale structural similarity for image quality assessment. In: 2003 The Thirty-Seventh Asilomar Conference on Signals, Systems and Computers, vol. 2, pp. 1398–1402. IEEE (2003)
15. Zubal, I.G., Harrell, C.R., Smith, E.O., Rattner, Z., Gindi, G., Hoffer, P.B.: Computerized three-dimensional segmented human anatomy. Med. Phys. **21**(2), 299–302 (1994)
16. Cheng, X., et al.: Direct parametric image reconstruction in reduced parameter space for rapid multi-tracer pet imaging. IEEE Trans. Med. Imaging **34**(7), 1498–1512 (2015)
17. Koeppe, R., Holthoff, V., Frey, K., Kilbourn, M., Kuhl, D.: Compartmental analysis of [11c] flumazenil kinetics for the estimation of ligand transport rate and receptor distribution using positron emission tomography. J. Cereb. Blood Flow Metab. **11**(5), 735–744 (1991)
18. Chen, S., Ho, C., Feng, D., Chi, Z.: Tracer kinetic modeling of/sup 11/c-acetate applied in the liver with positron emission tomography. IEEE Trans. Med. Imaging **23**(4), 426–432 (2004)
19. Terada, T., et al.: In vivo mitochondrial and glycolytic impairments in patients with Alzheimer disease. Neurology **94**(15), e1592–e1604 (2020)