Annals of Nuclear Medicine Online
Published online: 04 August 2021

Generative Image Transformer (GIT): unsupervised continuous image generative and transformable model for [^{123}I]FP-CIT SPECT images

Authors
Shogo Watanabe[1], Tomohiro Ueno[1], Yuichi Kimura[2], Masahiro Mishina[3], and Naozo Sugimoto[1]

Generative Image Transformer

1. Human Health Sciences, Graduate School of Medicine, Kyoto University, Japan.
2. Department of Computational Systems Biology, Faculty of Biology–Oriented Science and Technology, Kindai University, Wakayama, Japan.
3. Department of Neurology, Tokyo Rosai Hospital, Tokyo, Japan.

Mailing address: 53 Shogoin-Kawaharacho, Sakyo-ku, Kyoto City, Kyoto, Japan
TEL: 075-751-3934
Postal code: 606-8507
E-mail: wshogo1993@gmail.com
Type of article: Original article

Abstract

Objective: Recently, generative adversarial networks began to be actively studied in the field of medical imaging. These models are used for augmenting the variation of images to improve the accuracy of computer-aided diagnosis. In this paper, we propose an alternative new image generative model based on transformer decoder blocks and verify the performance of our model in generating SPECT images that have characteristics of Parkinson's disease patients.

Methods: Firstly, we proposed a new model architecture that is based on a transformer decoder block and is extended to generate slice images. From a few superior slices of 3D volume, our model generates the rest of the inferior slices sequentially. Our model was trained by using [^{123}I]FP-CIT SPECT images of Parkinson's disease patients that originated from the Parkinson's Progression Marker Initiative database. Pixel values of SPECT images were normalized by the specific/nonspecific binding ratio (SNBR). After training the model, we generated [^{123}I]FP-CIT SPECT images. The transformation of images of the healthy control case SPECT images into PD-like images was also performed. Generated images were visually inspected and evaluated using the mean absolute value and asymmetric index.

Results: Our model was successfully generated and transformed into PD-like SPECT images. The mean absolute SNBR was mostly less than 0.15 in absolute value. The variation of obtained dataset images was confirmed by the analysis of the asymmetric index.

Conclusions: These results showed the potential ability of our new generative approach for SPECT images that the generative model based on the transformer realized both generation and transformation by a single model.

Keywords: [^{123}I]FP-CIT SPECT, Generative model, Parkinson's disease, Transformer, Unsupervised learning

Introduction

In recent years, generative models such as the generative adversarial network (GAN) (1) are actively researched to generate medical images. Particularly, the deep convolutional GAN (DCGAN) (2) consists of convolutional neural network architectures that achieved good-quality images of computed tomography, magnetic resonance imaging (MRI), and positron emission tomography (PET).

Onishi et al. (3) reported that they augmented pulmonary nodule images generated by GAN and improved the classification performance by fine-tuning the pre-trained convolutional neural network (CNN). Koshino et al. (4) generated MRI images using the simple DCGAN approach. Islam et al. (5) generated normal control, mild cognitive impairment, and Alzheimer's disease PET images from random noise using DCGAN. Frid-Adar et al. (6) reported the result of improved liver lesion recognition performance with augmentation by GAN-based generated images.

Moreover, there are studies on pathological transformation by conditional GAN (7) or CycleGAN (8). Xia et al. (9) proposed the transformation of combined U-net (10) base encoder-decoder and conditional GAN to synthesize the aging variation brain images. Kimura et al. (11) demonstrated that CycleGAN could synthesize healthy controls to abnormal cases using unpaired PET images. Wei et al. (12) augmented minor case images using CycleGAN image translation. Liyan Sun et al. (13) proposed abnormal-to-normal translation GAN to synthesize MRI-image-contained lesions.

One of the motivations for generating or synthesizing medical images in previous studies is data augmentation for improving the performance of computer-aided diagnosis. We thought that it is important to have many kinds of generative models for ensuring varieties of augmented data. Therefore, in this paper, we propose an alternative new image generative model using Transformer (14).

The reason why we adopted Transformer in our generative model is that it is a powerful architecture for time series data. Our idea is that the slice order axis of 3D SPECT data can be treated as if it corresponds to the time axis. In that case, 3D SPECT data can be regarded as a time sequence, and we can use Transformer. We expect Transformer to be able to handle the continuity and the transition between slices. In the natural language processing community, Transformer has achieved state-of-the-art performance and demonstrated its significance on various tasks.

The number of studies on the application of Transformer for computer vision has gradually increased in recent years. In image captioning, Cornia et al. (15) proposed the Meshed-Memory Transformer, with learning connectivity between the encoder and the decoder, to enhance performance. Girdhar et al. (16) applied Transformer to a video context that understood and recognized actions. In object detection, Carion et al. (17) improved the end-to-end architecture and reduced hyperparameters like non-maximum suppression by the replaced inference design with the transformer encoder and decoder.

There are some movements to apply Transformer to image generation. Image Transformer (18) proposed a unique transformer architecture to predict pixels and to work on super-resolution. Image GPT (19) is an attempt to apply the GPT-2 (20) scheme to an unsupervised image generation. Image GPT consists of two stages: pre-training by an autoregressive model or BERT (21) and fine-tuning the model. It predicts next-to-next pixels of the subsequent pixel of the input pixels. These studies generate the subsequent parts of an input image by predicting pixels. Among studies based on Transformer architecture, there were no methods that directly generate whole images.

In this paper, we propose a new network model that consists of simple

Transformer decoder blocks to generate image slices. From a few superior slices of 3D volume data, our model generates the rest of inferior slices sequentially. We named the model Generative Image Transformer (GIT). To the best of our knowledge, in the medical imaging community, image generation research using Transformer is a new approach.

We demonstrate that our proposed model could generate Parkinson's disease SPECT images and transform healthy control SPECT images into images that are characteristic of Parkinson's disease.

Since GIT generates image slices sequentially, it can be applied to time series data. In our case, GIT predicts future scenes. There are some studies on predicting future scenes from past and current images in the field of video processing (22, 23). These approaches are composed of the image-specific CNN to feature extraction and recurrent neural network to predict time series data. They make the model more complicated and require more training time than Transformer. In contrast, GIT is composed of only transformer blocks and training speed is fast.

Materials and Methods

[^{123}I]FP-CIT SPECT images on Parkinson's Progressive Marker Initiative database

In this experiment, we used [^{123}I]FP-CIT SPECT from the Parkinson's Progressive Marker Initiative (PPMI) database (24). SPECT images in the PPMI database were normalized into a Montreal Neurological Institute (MNI) space.

We used 441 Parkinson's disease (PD) cases. All SPECT images measured $91 \times 109 \times 91$. We split 441 PD SPECT images to 391 training datasets and 50

validation datasets. We performed a 5-fold cross-validation and did not permit duplication in validation datasets.

We also used healthy control (HC) SPECT images from the PPMI database to verify the transformation ability of GIT.

Data preprocessing

For data augmentation, we applied left-right reflection on the axial plane and slight rotation (−3 and 3 degree) on the axial, sagittal, and coronal planes to 391 training datasets. After augmentation, the number of training datasets changed to 5,474. All voxels of [^{123}I]FP-CIT SPECT images were normalized by the specific/nonspecific binding ratio (SNBR) (25) like formula,

$$SNBR = \frac{C - C_r}{C_r} \quad (1)$$

where C and C_r denote the concentration at each voxel and the mean concentration of the reference region, respectively. We calculated the mean value of the whole brain region, except the area around the striatum, as the reference region.

Network Architecture

Figure 1 shows the overview of our model's architecture. We only used Transformer decoder to build the network for the training autoregression model. At first, the input data transformed 1024 hidden feature vectors by position-wise fully connection. Then, we added positional encoding to each hidden feature vector to support slice order information. We used sinusoidal positional encoding following the original transformer.

We constructed a pre-layer normalization transformer architecture (26) instead of an original post-layer normalization one. In each transformer decoder block, layer normalization (27), multi-head attention, and residual connection were applied to the

hidden feature vectors. Thereafter, hidden feature vectors were applied during pre-layer normalization and the position-wise feedforward phase. We used mish (28) activation function on the position-wise feedforward phase to smoothly optimize the loss function. We inserted dropout (29), with dropout rate of 0.1 after multi-head attention and intermediate of position-wise feedforward phase for preventing overfitting.

In our experiment, we built a 16-layer transformer decoder block. The final output directly predicts the SNBR value on the next slice by fully connecting with the ELU (30) activation function, because the SNBR value is a real number with a lower limit of −1. The number of our model parameters is approximately 170 million.

Training and Inference

We defined t as the slice number. Figure 2 shows the training scheme in an autoregressive manner. To avoid getting information of future slices, we used masked self-attention in transformer decoder block. Therefore, the GIT is a unidirectional model and can only use information until the current input.

In this study, we used a regression approach to directly predict the SNBR value and simply minimized the sum squared loss as follows:

$$loss = \sum_{t=1}^{T-1} \sum_{v=1}^{V} \left(\boldsymbol{x}_{t+1,v} - \hat{\boldsymbol{x}}_{t+1,v} \right)^2 \tag{2}$$

$$\hat{\boldsymbol{x}}_{t+1,v} = Trm\left(\boldsymbol{x}_{t,v} \right) \tag{3}$$

where T denotes total number of image slices and V denotes the total number of voxels on each image slice. Trm represents our model, which predicts voxels of the next image slice.

We chose the Adam optimizer (31) with its parameter $\beta_1 = 0.9, \beta_2 = 0.99$ and used the Cyclical Learning Rate (CLR) (32) for learning rate scheduling. We set the base learning rate to 1.0×10^{-5}, maximum learning rate to 1.0×10^{-3}, and

triangular2 policy.

We trained our model using 5,474 training datasets. We trained by minibatch training with 10 minibatch sizes and 100 epochs. These parameters were defined by trial and error. To avoid underfitting or overfitting, a learning curve of a loss function and generated images were observed. Observed images under smaller epoch time were mostly blurred.

At inference, our model repeats to generate the $t+1$th image slice from t image slices until the maximum number of slices.

We used the Microsoft Cognitive Toolkit (CNTK) (33) deep learning framework. In our implementation, we ran the training model on NVIDIA GPU Quadro RTX 6000 24GB.

Validation

For evaluating the model, we generated SPECT images from validation data and HC data. From superior 40 slice images, the rest of 51 inferior slices were generated. Image generation from fewer input slices was also attempted. From the superior 15 slices, the rest of the 76 slices were generated.

We evaluated our generated images qualitatively and quantitatively.

In qualitative evaluations, visual inspections of generated images were performed.

In quantitative evaluations, we calculated mean absolute value of 51 image slices between each validation dataset and the generated SPECT. We also evaluated the variation of the generated images by using the asymmetric index (AI) explained in the following. In the clinical diagnosis of PD, the striatum is the most important region on the SPECT image. We extracted the striatal voxel-of-interest (VOI), whose shape of height and width is 60 × 90 mm and 11 slices. We calculated the AI (34) using the

following equation:

$$\mathrm{AI} = \frac{SNBR_{left} - SNBR_{right}}{SNBR_{left} + SNBR_{right}} \quad (4)$$

where $SNBR_{left}$ and $SNBR_{right}$ are the striatum uptake count on the left and right striatum in VOI, respectively. The original formula of AI uses the absolute value; however, we defined AI with a signed value for obtaining the details of the left-right spatial information.

Results

Figure 3 shows the generated image from the 40 slices to the subsequent 51 slices. The left and right images are the original and generated image, respectively. The red rectangle includes the generated image slices by our model. The color maps of original and generated images have the same range of values. For qualitative evaluation, the first author reviewed all generated images and confirmed that [^{123}I]FP-CIT SPECT-like images were successfully generated in almost cases. The variations of the tissue and the shape of generated images were observed. In addition, three clinicians (the fourth author: neurologist, another neurologist, and radiologist) reviewed randomly sampled 20% cases in fold #2 and confirmed that generated images mostly had features of PD, which are a declining SNBR value, an asymmetry, and the shape of the striatum. No unnatural discontinuity was observed in axial images.

Figure 4 shows three cross-sectional images of a generated sample. Left top, right top, and right bottom represent sagittal, coronal, and axial image slices, respectively. Although the GIT predicts axial images in a slice-by-slice manner, no unnatural discontinuity between adjacent generated axial slices was observed in the sagittal and coronal planes. However, a slight discontinuity was observed between the last input slice (40th slice) and the first generated slice (41st slice).

Figure 5 shows the input and the generated images of the HC case in the PPMI database. The red rectangle includes 40 transformed images. The value of SNBR declined and the difference between the left and right striatum became apparent.

Figure 6 shows the mean absolute value per voxels in each slice among validation datasets and SPECT images generated by our model in 5-fold. The horizontal and vertical axis represents the slice number and the mean absolute value per slice, respectively. The center dot represents the mean and the error bar represents the standard deviation. The mean and standard deviation of the mean absolute value are less than approximately 0.15 in the slices from 41 to 55 that almost covers the striatum region. After 77 slices, the error and the standard deviation get bigger. The following approximately 80 slices are not important because of the outside of the brain.

Figure 7 shows the mean absolute value map between validation and generated images of fold #2. The horizontal axis represents slice number and the vertical axis represents case number in fold #2. Each pixel represents the mean value of absolute error of the slice on the case. We sorted the cases following the absolute error. The right bottom of the map has a larger error than other regions. The error value of the blue region is less than 0.15, and we regarded it as the successful case. Case 39, in the bottom row, is the worst case. Figure 8 shows estimated slices of the worst case in Fig. 7. In that case, however, the mean absolute error was small enough until about the 55th slice.

To investigate the variation of the generated images, we compared the SNBR value difference between the left and right striatum based on AI. Figure 9 shows the histogram of AI of validation datasets and generated SPECT images in 5-fold. Negative and positive values of AI mean declines of the right and left striatum, respectively. The

histogram of AI of validation datasets is characterized by the right striatum decline being bigger than the left one. The histogram of AI of the generated images has a smaller width compared to validation images. However, our model could reproduce the feature of original datasets, wherein the right striatum decline is larger than the left one.

Figure 10 shows the generated image from the first 15 slices to the rest of the 76 slices. The red region includes the generated image slices. These validation data are the same as those in Fig. 3. GIT could generate the remaining image slices from only a few image slices. However, the SPECT images generated from 15 slices are a little blurrier than those generated from 40 slices.

Discussion

We proposed a new generative image model approach by transformer architecture and an autoregressive unsupervised training scheme. In this paper, our model has demonstrated that a single model could generate the rest of an image and could transform the pathological features of an image.

We proposed a new generative image model approach by transformer architecture and autoregressive unsupervised training scheme. In this paper, our model has demonstrated that it could generate the rest of images, including its pathological features.

We presented that our model could generate the rest of slices from input slices, and the generated images are confirmed realistic SPECT images by experts (Figs. 3, 4, 5, and 10).

In addition, our model trained in this paper can generate SPECT images, which have the features of PD. We showed that the generated SPECT images had asymmetric indices from −0.25 to 0.2 (0.45 in the total range) as shown in Fig. 9, which was

comparable to those of validation images from −0.35 to 0.15 (0.5 in the total range). In addition, we demonstrated that our model not only extrapolates the PD images but also generates PD-like images from the HC SPECT image, which were not included in the training data at all (Fig. 5). In the generation of PD-like images from HC images, our model created the features of the PD image, the declined SNBR value. The generation of PD-like images from HC images does not imply diagnostic prediction. Our present model just generates images only but does not make a diagnostic prediction as to whether the HC case will become PD or not. Mechanisms in the generation of asymmetry from symmetric input data should be investigated further in our future work.

We also demonstrated that GIT could possibly generate images from only 15 input image slices (Fig. 10). Although PD-like images were also obtained, image resolution was slightly lower than images generated from a 40-slice input. Lower resolution is thought to be caused by regression into an average image because the input information is less.

In our approach, we can prepare paired targets and predicted data from only the validation dataset. That allows quantitative validation to measure the precision of the model. This is one of the advantages of our approach.

Our model could precisely predict approximately 15 image slices (41st~55th) (Fig. 6, 7). The error is small until about the 55th slice. Although the standard deviation increases from about the 70th, these image slices are not important for diagnosis (whether PD or not) because these slices are out of the brain region.

In the field of image generation, we are interested in the variation of generated images. Although the generative model generally produces average images, it is not acceptable. Therefore, we explored that our model had variations of asymmetric

striatum, which is important to diagnose PD on generated images (Fig. 9). In the training phase, we adopted left-right reflection for data augmentation. Then, the average image slices of training data were completely symmetric. Although the variation of our model is a little smaller than the validation datasets, our model could produce enough variation of the striatum. The decline of variations could be caused by sum squared loss in the training phase because L2 loss assumed the normal distribution. In future work, we need to consider the more effective loss function.

Our model sometimes failed to generate images in the second half. GIT, however, could generate more than 15 image slices that cover the striatum region even in the worst case (Fig. 8).

The limitation of GIT is the need for a few input image slices. Therefore, GIT does not support the generation of data from latent vectors of random noise like GAN. However, GIT can generate PD images and generate PD-like images from HC images in the single model.

Other limitations of the present study originated from the use of the PPMI database. Since SPECT images in the PPMI database were normalized into the MNI space, image interpolation degraded the image resolution. Also, the images we used have a lower resolution than modern realistic SPECT or PET images. Although the lower resolution might reduce the degree of variation in a disease representation, the PPMI database has high accessibility and reliability. To evaluate the feasibility of the use of our model in image generation, especially in the first step, accessibility and reliability seem important. In this regard, we only used the PPMI database to verify image generation by our model in this paper. Many atypical PD cases are encountered in the realistic clinical situations, PPMI database; however, this includes just regular

staged PD data. Therefore, much higher degrees of variation will be expected in clinical situations. In a future study, we should address whether our model can maintain a higher resolution or higher degree of variation even in those realistic clinical situations. We expect that the current model could be extended by using such techniques as a transfer learning technique. Since higher-resolution images may have a rich information in the superior slices, it can be useful for GIT to generate inferior slices.

Acknowledgments

The data we used in training and validation in this study were obtained from the Parkinson's Progression Marker Initiative (PPMI) database (https://www.ppmi-info.org/data). PPMI—a public private partnership—was funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including Abbvie, Avid, Biogen Idec, Bristol-Myers Squibb, Covance, GE Healthcare, Genentech, GlaxoSmithKline, Lilly, Lundbeck, Merck, Meso Scale Discovery, Pfizer, Piramal, Roche, and UCB.

The authors thank professor Nobukatsu Sawamoto (MD, neurologist) and Associate professor Koichi Ishizu (MD, radiologist) for reviewing generated SPECT images. Both belong to the Graduate School of Medicine, Kyoto University.

References

1. Goodfellow Ian J, Pouget-Abadie J, Mirza M, Xu Bing, Warde-Farley D, Ozair S, et al. Generative adversarial nets. AdvNeural Inf Process Syst2014, p. 2672–2680.
2. Radford A, Metz L, and Chintal S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. arXiv preprint arXiv:1511.06434 (2015).
3. Onishi Y, Teramoto A, Tsujimoto M, Tsukamoto T, Saito K, Toyama H, et al., Automated pulmonary nodule classification in computed tomography images using a deep convolutional neural network trained by generative adversarial networks. BioMed Red. Int. 2019.
4. Koshino K, Werner Rudolf A, Toriumi F, Javadi Mehrbod S, Pomper Martin G, Solnes Lilja B, et al. Generative adversarial networks for the creation of realistic artificial brain magnetic resonance images. Tomography, 2018. vol. 4, no. 4, 159.
5. Islam J and Zhang Y. GAN-based synthetic brain PET image generation. Brain Informatics, 2020. vol. 7, p. 1–12.
6. Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J, and Greenspan H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing, 2018. vol. 321, pp. 321–331.
7. Mirza M and Osindero S. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784 (2014).
8. Zhu J-Y, Park T, Isola P, and Efros A A. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. the IEEE international conference on computer vision, 2017. p. 2223–2232.
9. Xia T, Chartsias A, and Tsaftaris S A. Consistent brain ageing synthesis, in Med. Image Comput Assist. Interv. Springer, Cham, 2019. p. 750–758.

10. Ronneberger O, Fischer P, and Brox T. U-Net: convolutional networks for biomedical image segmentation. Med. Image Comput Assist. Interv. Springer, Cham, 2015. p. 234–241.
11. Kimura Y, Watanabe A, Yamada T, Watanabe S, Nagaoka T, Nemoto M, et al. AI approach of cycle-consistent generative adversarial networks to synthesize PET images to train computer-aided diagnosis algorithm for dementia. Ann. Nucl. Med. 2020. p. 1–4.
12. Wei J, Suriawinata A, Vaickus L, Ren Bing, Liu X, Wei J, et al. Generative Image Translation for Data Augmentation in Colorectal Histopathology Images. arXiv preprint arXiv:1910.05827 (2019).
13. Liyan Sun, Wang J, Huang Yue, Ding X, Greenspan H, and Paisley J. An adversarial learning approach to medical image synthesis for lesion detection. IEEE J. Biomed. Health Inform, 2020. p. 2303–2314.
14. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, et al. Attention is all you need. In Adv. Neural Inf. Process. Syst. 2017. p. 5998–6008.
15. Cornia M, Stefanini M, Baraldi L, and Cucchiara R. Meshed-memory transformer for image captioning. In Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2020. p. 10578–10587.
16. Girdhar R, Carreira J, Doersch C, and Zisserman A. Video action transformer network. In Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2019. p. 244–253.
17. Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A, and Zagoruyko S. End-to-end object detection with transformers. arXiv preprint arXiv:2005.12872 (2020).

18. Parmar N, Vaswani A, Uszkoreit J, Kaiser Ł, Shazeer N, Ku A, et al. Image transformer. arXiv:1802.05751 (2018).
19. Chen M, Radford A, Child R, Wu J, Jun H, Luan D, et al. Generative pretraining from pixels. In: Proceedings of the 37th International Conference on Machine Learning. 2020.
20. Radford A, Wu J, Child R, Luan D, Amodei D, and Sutskever I. Language models are unsupervised multitask learners. OpenAI Blog, 2019. vol 1, no. 8, p 9
21. Devlin J, Chang M-W, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018).
22. Fragkiadaki K, Agrawal P, Levine S, and Malik J. Learning visual predictive models of physics for playing billiards. arXiv preprint arXiv:1511.07404 (2015).
23. Lotter W, Kreiman G, and Cox D. Deep predictive coding networks for video prediction and unsupervised learning. arXiv preprint arXiv:1605.08104 (2016).
24. Marek K, Jennings D, Lasch S, Siderowf A, Tanner C, Simuni T, et al. The parkinson progression marker initiative (PPMI). Prog. Neurobiol. 2011. vol. 95, p 629–35.
25. Tossici-Bolt L, Hoffmann S M A, Kemp P M, Mehta R L, Fleming J S. Quantification of [^{123}I]FP–CIT SPECT brain images: an accurate technique for measurement of the specific binding ratio. Eur. Img. J. Nucl. Med. Mol. 2006. vol. 33, p 1491–1499.
26. Xiong R, Yang Y, He D, Zheng K, Zheng S, Xing C, et al. On layer normalization in the transformer architecture. arXiv preprint arXiv:2002.04745 (2020).

27. Ba J L, Kiros J R, and Hinton G E. Layer normalization. arXiv preprint arXiv:1607.06450 (2016).
28. Diganta M. Mish: A self regularized non-monotonic neural activation function. arXiv preprint arXiv:1908.08681 (2019).
29. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn Res. 2014. vol. 15, no. 1, p 1929–1958.
30. Clevert D-A, Unterthiner T, and Hochreiter S. Fast and accurate deep network learning by exponential linear units (ELUs). arXiv preprint arXiv:1511.07289 (2015).
31. Kingma D P and Ba J L. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014).
32. Smith L N. Cyclical learning rates for training neural networks. In IEEE Winter Conference on Applications of Computer Vision. 2017. p. 464–472.
33. Seide F and Agarwal A. CNTK: Microsoft's open-source deep-learning toolkit. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016. p. 2135–2135.
34. Hayashi T, Mishina M, Sakamaki M, Sakamoto Y, Suda S, Kimura K. Effect of brain atrophy in quantitative analysis of 123I iofupane SPECT. Ann. Nucl. Med. 2019. vol 33, no. 8, pp 579–585.

Figures

Fig. 1

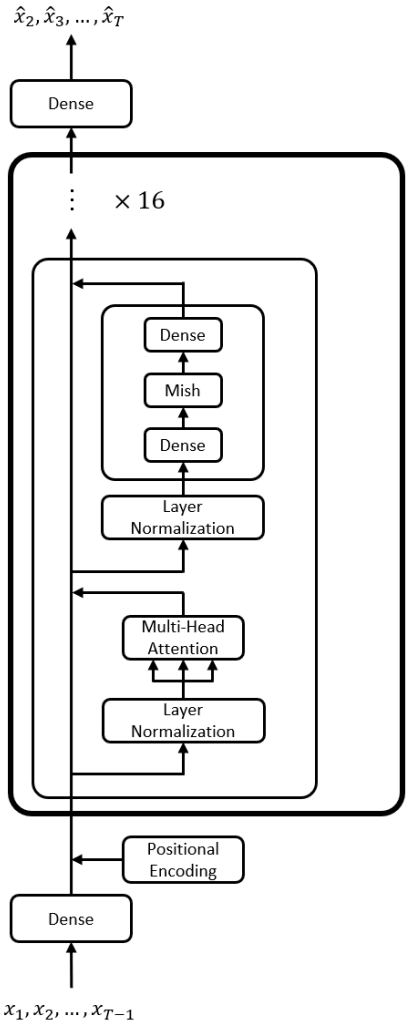

Fig. 1. Our implemented transformer decoder block. At first, input data are applied position-wise fully connected to embed hidden dimensional features. Next, sinusoidal positional encodings are added. Transformer decoder blocks are stacked in 16 layers. We constructed pre-layer normalization to multi-head masked self-attention and to position-wise feedforward. Finally, features are applied layer normalization and fully connected with ELU activation to predict SNBR values directly

Fig. 2

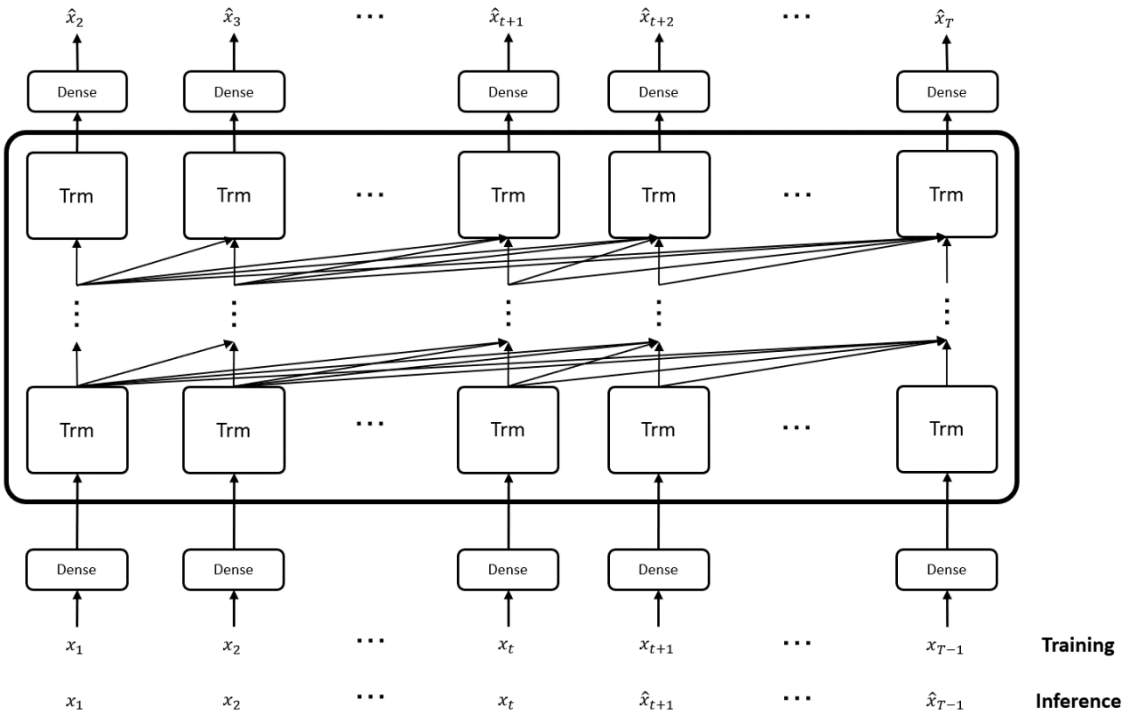

Fig. 2. Our model training scheme operates in an autoregressive manner like OpenAI GPT (18). Trm represents our transformer decoder block. The model predicts next input $\hat{x}_2$ based on up to current input x_1. Therefore, the target data of $\hat{x}_2$ is original next input x_2. When we train the model, we can simultaneously feed all slices to the model because masked self-attention for future information applies in transformer decoder block. At inference, the model needs some image slices until t and predicts the t+1th image slice

Fig. 3

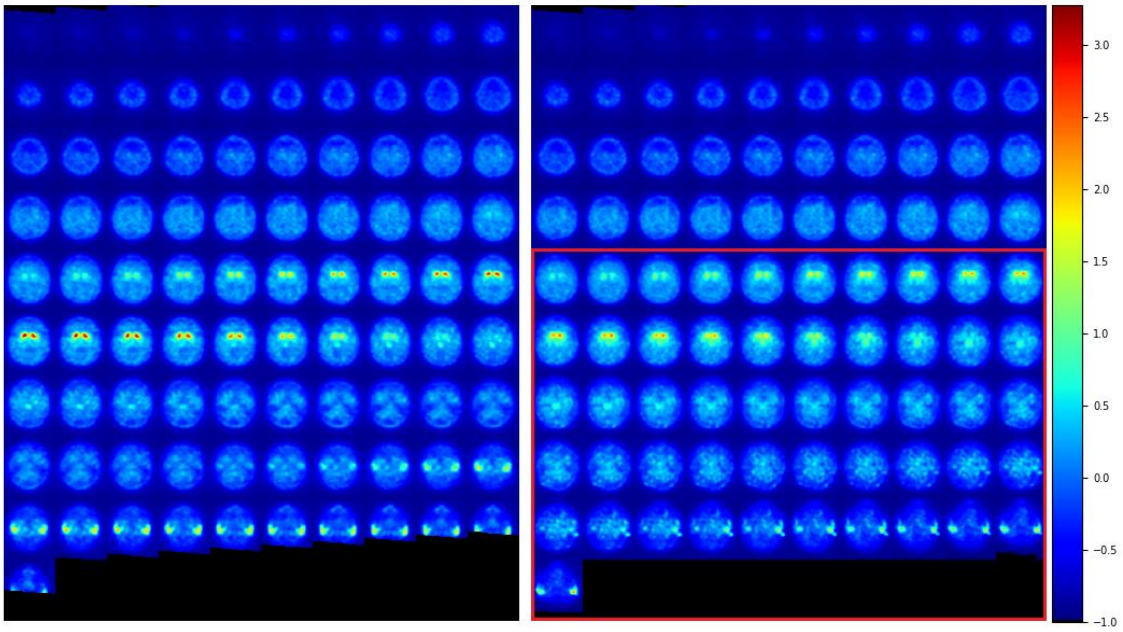

Fig. 3. The input data are 40 image slices from the original SPECT image (left). Output data predicted that our model is the subsequent 51 image slices. The region surrounded by the red rectangle shows the generated image slices

Fig. 4

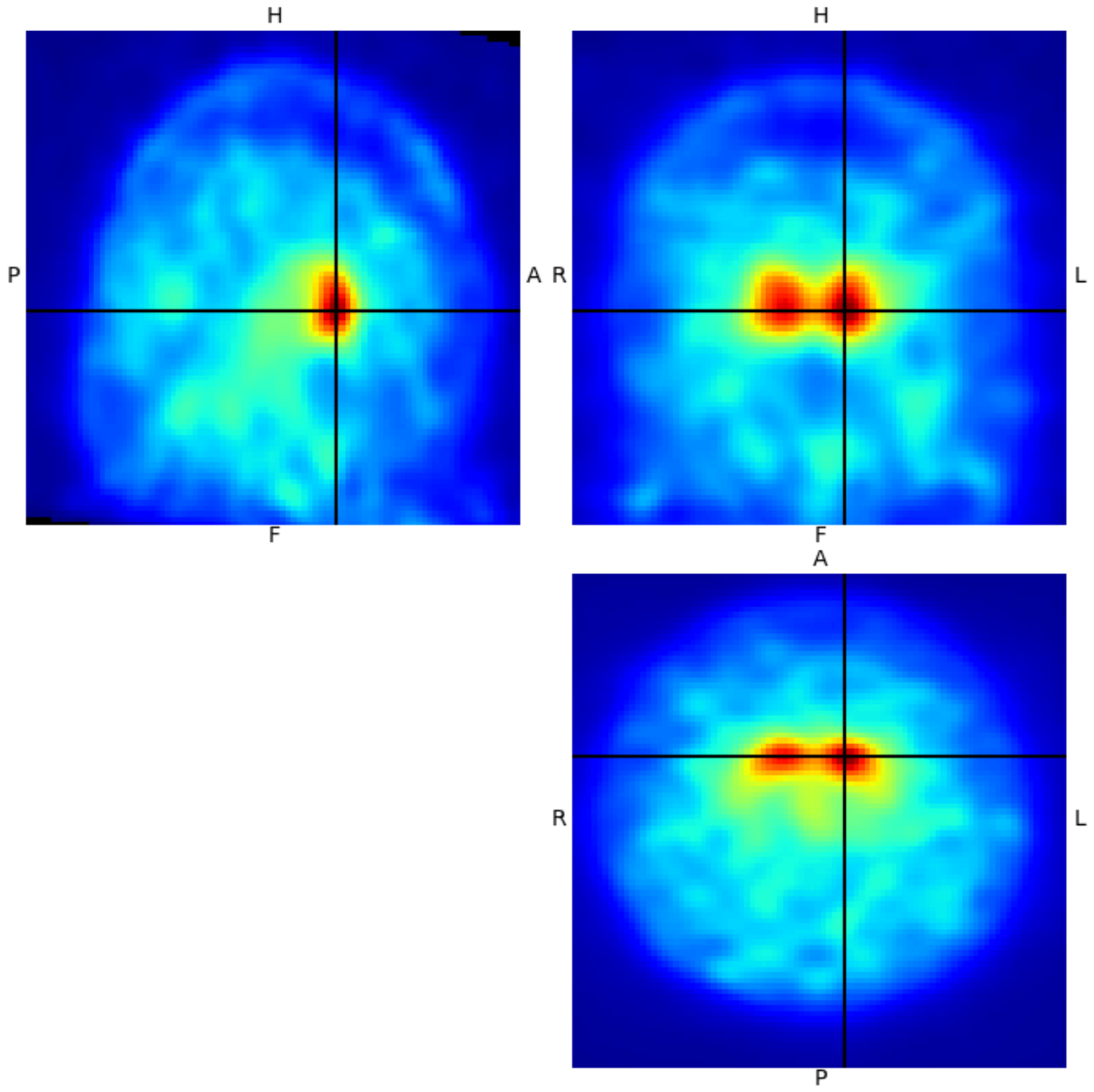

Fig. 4. We displayed generated data from the 40 slices of validation data as 3D tomography. Left top, right top, and right bottom show the sagittal, coronal, and axial image slice, respectively. Our model only used information about the axial image slice; therefore, the gap between input data and generated data in the coronal and sagittal directions

Fig. 5

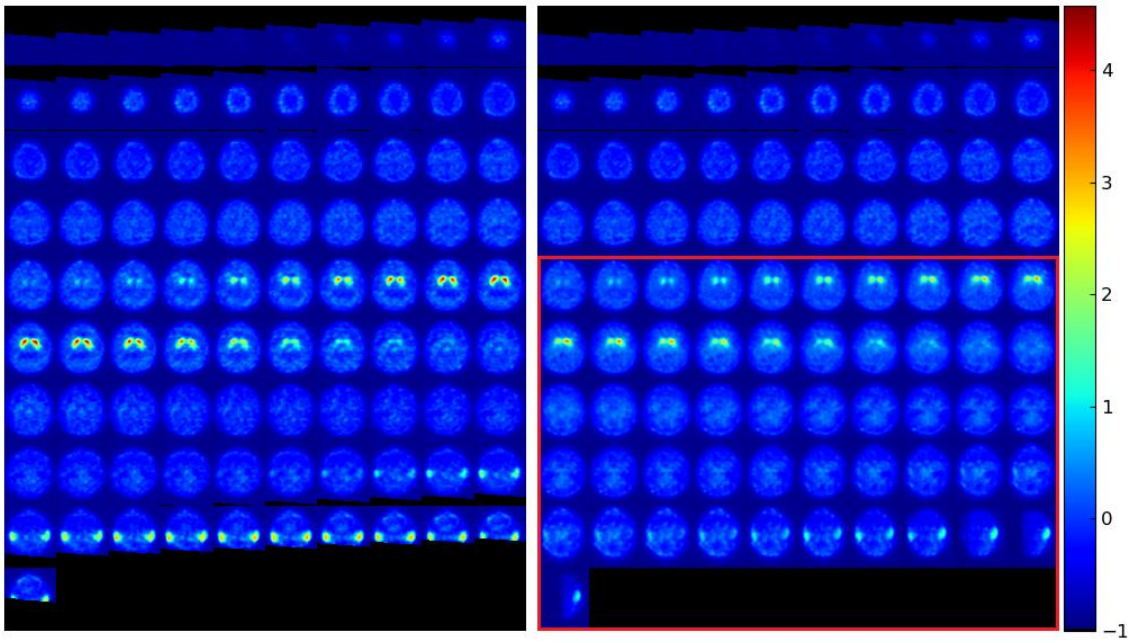

Fig. 5. An example of a transformation from a healthy control case to a PD case. The left image is original healthy control SPECT images. The right one is the generated image transformed by the subsequent of healthy control from the 40th slice. The small SNBR value in the striata, especially the right striatum, is declined

Fig. 6

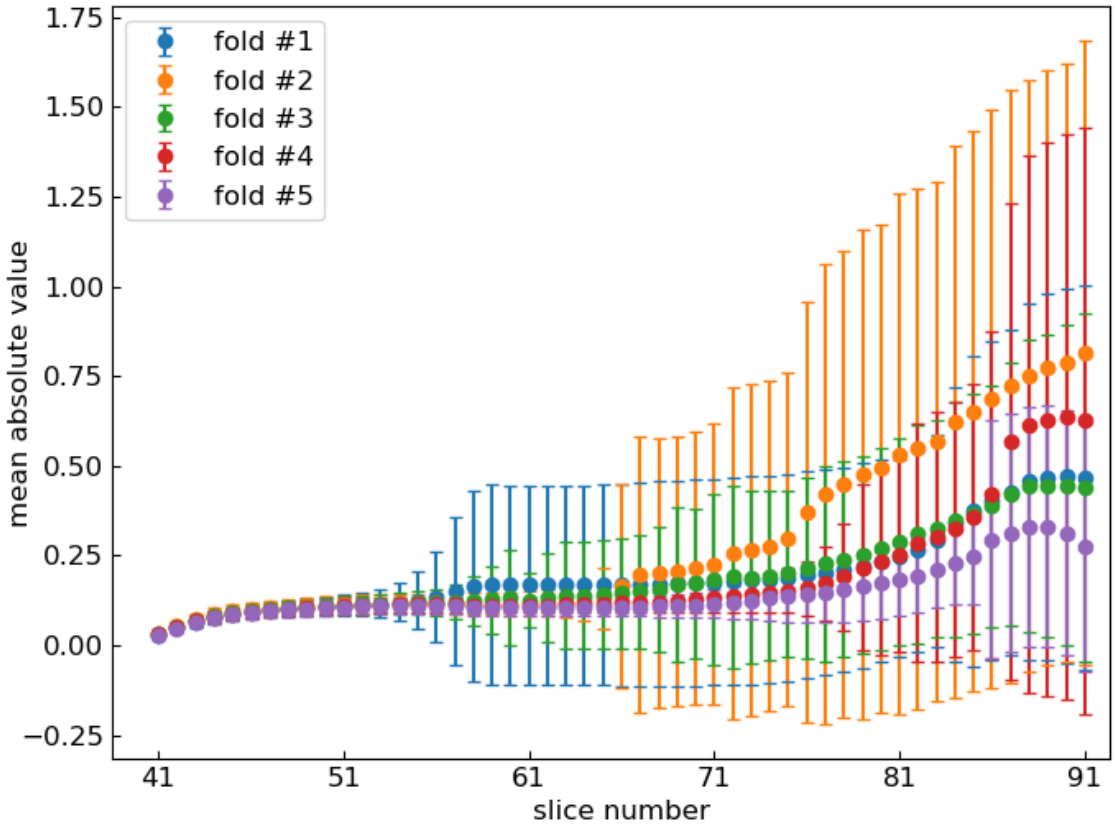

Fig. 6. The mean and standard deviation of the mean absolute error between validation and generated images in 5-fold. The horizontal axis represents the slice number from 41 to 91. The vertical axis represents the mean absolute error. The center dot represents the mean and error bars represent the standard deviation

Fig. 7

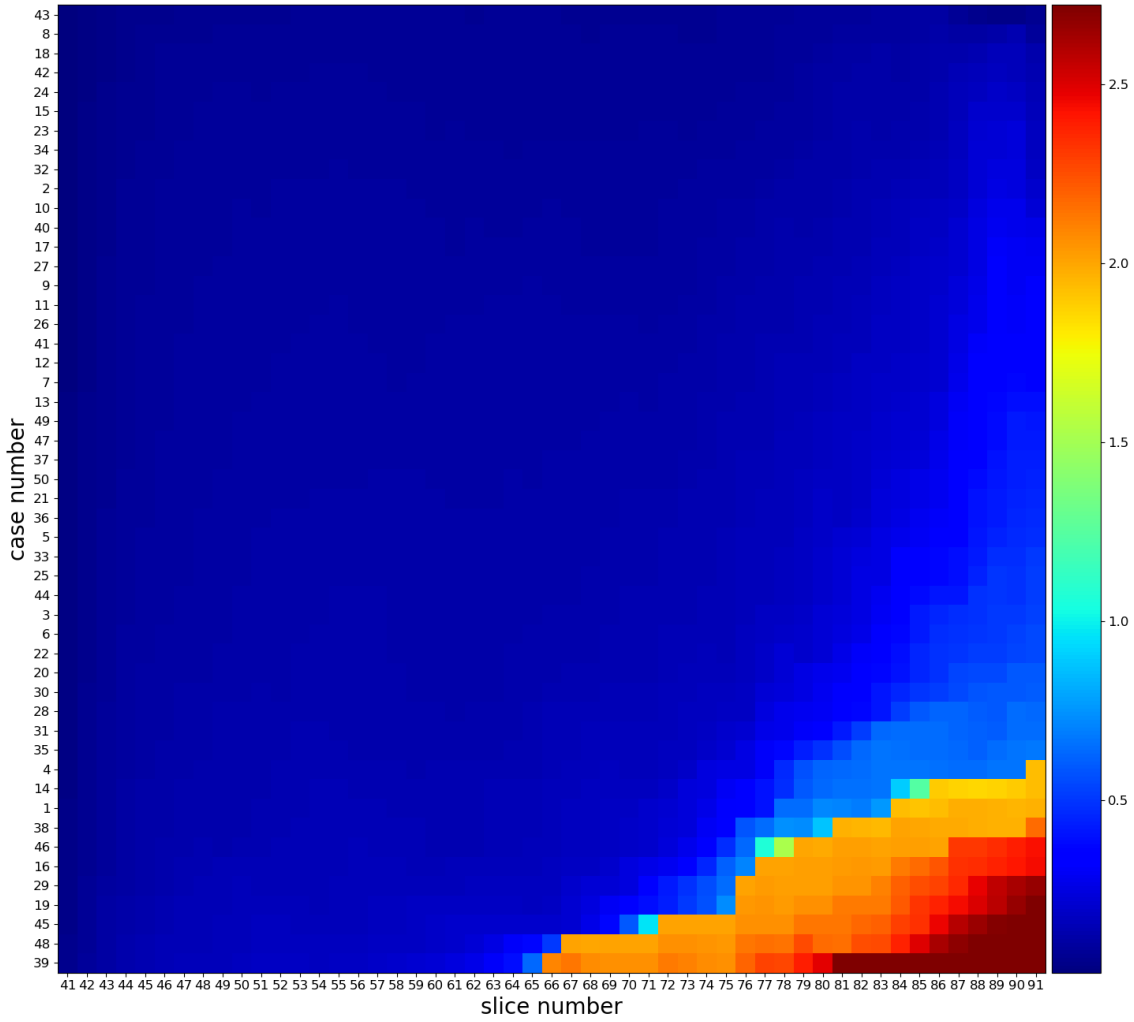

Fig. 7. The mean absolute error map of fold #2. The horizontal axis represents the slice number and the vertical axis represents the case number. Each pixel represents the mean absolute error in each slice on each case, and we sorted in ascending order based on the mean absolute error of each case. The right bottom of the map is the largest error case

Fig. 8

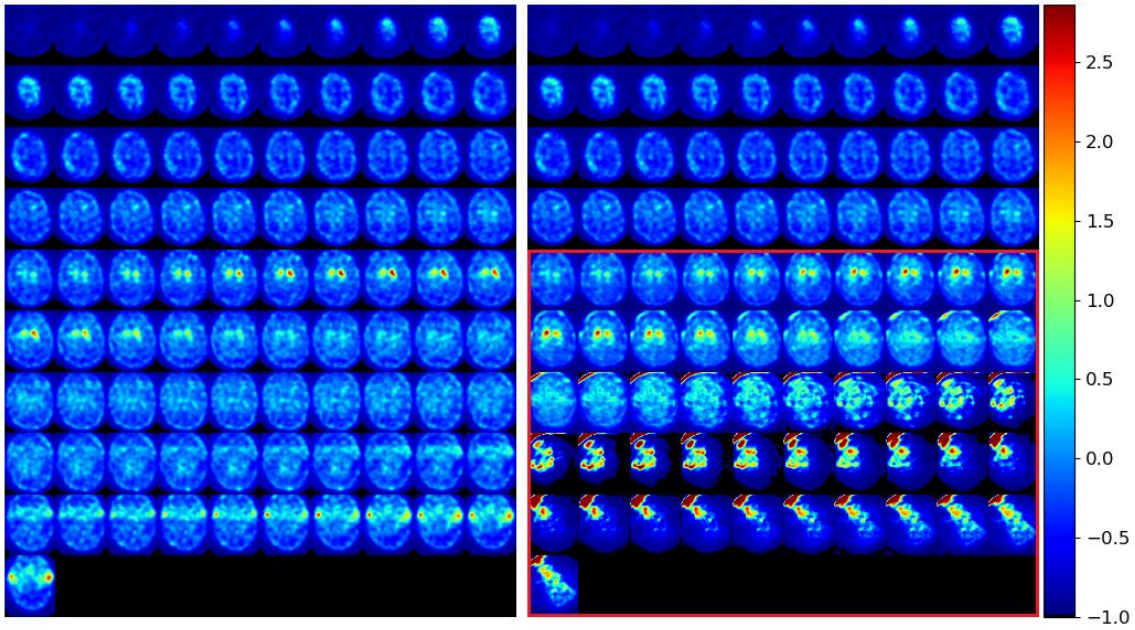

Fig. 8. The worst case in fold #2. The region surrounded by the red rectangle shows the generated image slices. GIT could succeed up to approximately 15 image slices

Fig. 9

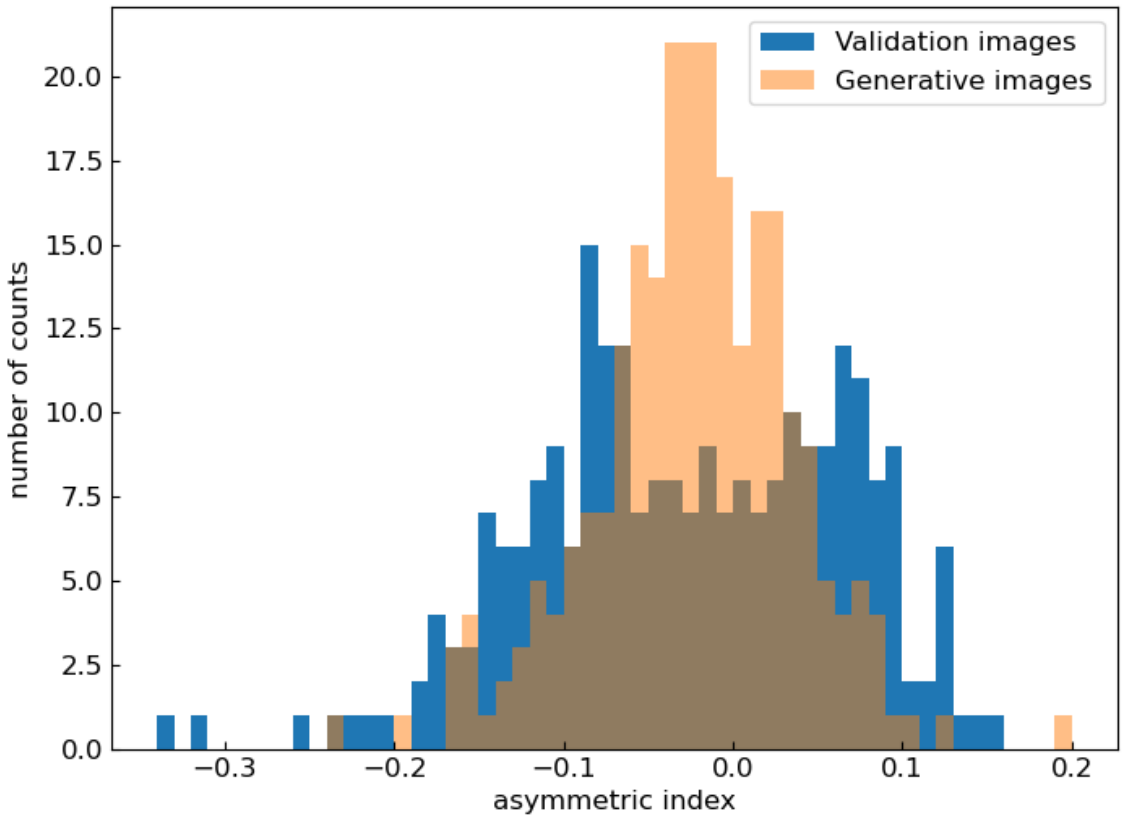

Fig. 9. Blue histogram is AI of validation datasets and orange one is the generated images on 5-fold. The horizontal axis represents the asymmetric index and the vertical axis represents the number of counts. The negative AI represents the right striatum decline, the 0 represents no difference between left and right, and the positive represents left striatum decline

Fig. 10

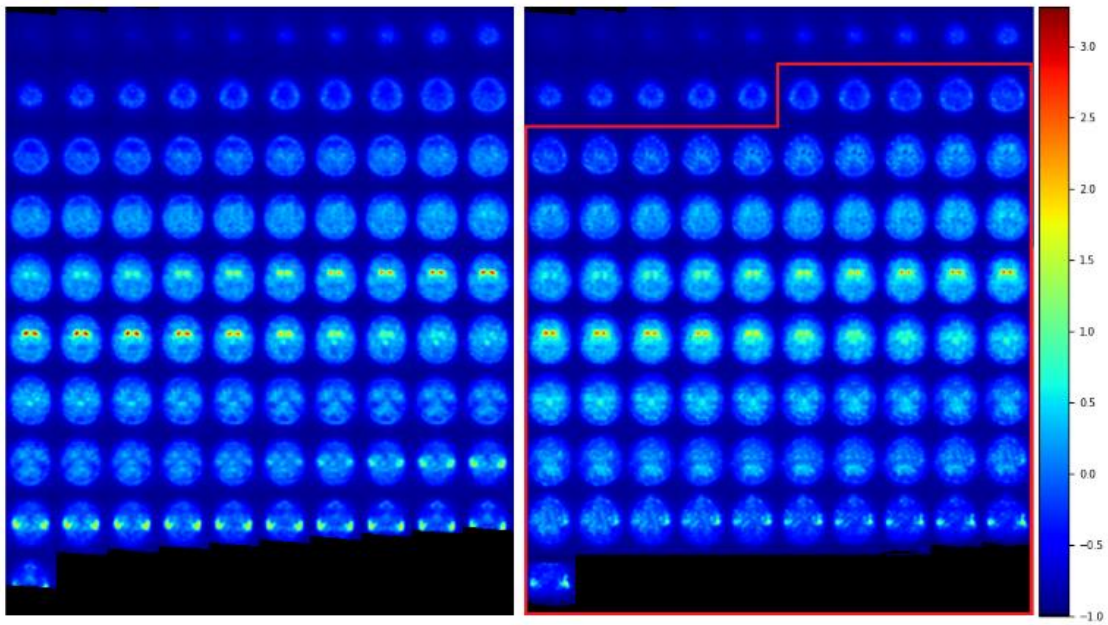

Fig. 10. Input data are 15 image slices from an original SPECT image (left). Output data predicted that our model is the subsequent 76 image slices. The region surrounded by the red rectangle shows generated image slices

Acknowledgments

This is a post-peer-review, pre-copyedit version of an article published in Annals of Nuclear Medicine. The final authenticated version is available online at: https://doi.org/10.1007/s12149-021-01661-0.