DRMC: A Generalist Model with Dynamic Routing for Multi-center PET Image Synthesis

Zhiwen Yang[1], Yang Zhou[1], Hui Zhang[2], Bingzheng Wei[3], Yubo Fan[1], and Yan Xu[1](✉)

[1] School of Biological Science and Medical Engineering, State Key Laboratory of Software Development Environment, Key Laboratory of Biomechanics and Mechanobiology of Ministry of Education, Beijing Advanced Innovation Center for Biomedical Engineering, Beihang University, Beijing 100191, China
xuyan04@gmail.com
[2] Department of Biomedical Engineering, Tsinghua University, Beijing 100084, China
[3] Xiaomi Corporation, Beijing 100085, China

Abstract. Multi-center positron emission tomography (PET) image synthesis aims at recovering low-dose PET images from multiple different centers. The generalizability of existing methods can still be suboptimal for a multi-center study due to domain shifts, which result from non-identical data distribution among centers with different imaging systems/protocols. While some approaches address domain shifts by training specialized models for each center, they are parameter inefficient and do not well exploit the shared knowledge across centers. To address this, we develop a generalist model that shares architecture and parameters across centers to utilize the shared knowledge. However, the generalist model can suffer from the center interference issue, *i.e.* the gradient directions of different centers can be inconsistent or even opposite owing to the non-identical data distribution. To mitigate such interference, we introduce a novel dynamic routing strategy with cross-layer connections that routes data from different centers to different experts. Experiments show that our generalist model with dynamic routing (DRMC) exhibits excellent generalizability across centers. Code and data are available at: https://github.com/Yaziwel/Multi-Center-PET-Image-Synthesis.

Keywords: Multi-Center · Positron Emission Tomography · Synthesis · Generalist Model · Dynamic Routing

1 Introduction

Positron emission tomography (PET) image synthesis [1–10] aims at recovering high-quality full-dose PET images from low-dose ones. Despite great success,

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43898-1_4.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14222, pp. 36–46, 2023.
https://doi.org/10.1007/978-3-031-43898-1_4

most algorithms [1,2,4,5,8–10] are specialized for PET data from a single center with a fixed imaging system/protocol. This poses a significant problem for practical applications, which are not usually restricted to any one of the centers. Towards filling this gap, in this paper, we focus on multi-center PET image synthesis, aiming at processing data from multiple different centers.

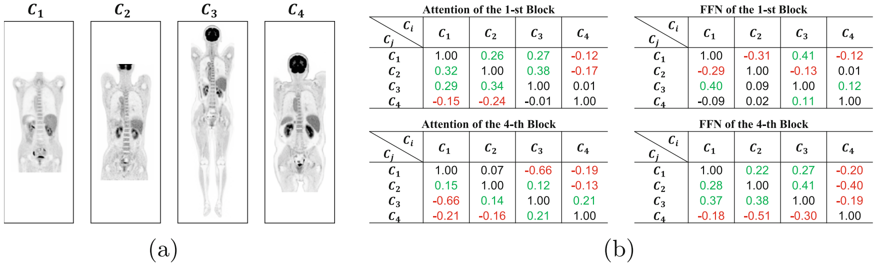
However, the generalizability of existing models can still be suboptimal for a multi-center study due to domain shift, which results from non-identical data distribution among centers with different imaging systems/protocols (see Fig. 1 (a)). Though some studies have shown that a specialized model (*i.e.* a convolutional neural network (CNN) [3,6] or Transformer [9] trained on a single center) exhibits certain robustness to different tracer types [9], different tracer doses [3], or even different centers [6], such generalizability of a center-specific knowledge is only applicable to small domain shifts. It will suffer a severe performance drop when exposed to new centers with large domain shifts [11]. There are also some federated learning (FL) based [7,11,12] medical image synthesis methods that improve generalizability by collaboratively learning a shared global model across centers. Especially, federated transfer learning (FTL) [7] first successfully applies FL to PET image synthesis in a multiple-dose setting. Since the resultant shared model of the basic FL method [12] ignores center specificity and thus cannot handle centers with large domain shifts, FTL addresses this by finetuning the shared model for each center/dose. However, FTL only focuses on different doses and does not really address the multi-center problem. Furthermore, it still requires a specialized model for each center/dose, which ignores potentially transferable shared knowledge across centers and scales up the overall model size.

A recent trend, known as generalist models, is to request that a single unified model works for multiple tasks/domains, and even express generalizability to novel tasks/domains. By sharing architecture and parameters, generalist models can better utilize shared transferable knowledge across tasks/domains. Some pioneers [13–17] have realized competitive performance on various high-level vision tasks like classification [13,16], object detection [14], *etc.*

Nonetheless, recent studies [16,18] report that conventional generalist [15] models may suffer from the interference issue, *i.e.* different tasks with shared parameters potentially conflict with each other in the update directions of the gradient. Specific to PET image synthesis, due to the non-identical data distribution across centers, we also observe the **center interference issue** that the gradient directions of different centers may be inconsistent or even opposite (see Fig. 1). This will lead to an uncertain update direction that deviates from the optimal, resulting in sub-optimal performance of the model. To address the interference issue, recent generalist models [14,16] have introduced dynamic routing [19] which learns to activate experts (*i.e.* sub-networks) dynamically. The input feature will be routed to different selected experts accordingly so as to avoid interference. Meanwhile, different inputs can share some experts, thus maintaining collaboration across domains. In the inference time, the model can reasonably generalize to different domains, even unknown domains, by utilizing the knowledge of existing experts. In spite of great success, the study of generalist models rarely targets the problem of multi-center PET image synthesis.

In this paper, inspired by the aforementioned studies, we innovatively propose a generalist model with **D**ynamic **R**outing for **M**ulti-**C**enter PET image synthesis, termed DRMC. To mitigate the center interference issue, we propose a novel dynamic routing strategy to route data from different centers to different experts. Compared with existing routing strategies, our strategy makes an improvement by building cross-layer connections for more accurate expert decisions. Extensive experiments show that DRMC achieves the best generalizability on both known and unknown centers. Our contribution can be summarized as:

- A generalist model called DRMC is proposed, which enables multi-center PET image synthesis with a single unified model.
- A novel dynamic routing strategy with cross-layer connection is proposed to address the center interference issue. It is realized by dynamically routing data from different centers to different experts.
- Extensive experiments show that DRMC exhibits excellent generalizability over multiple different centers.



Attention of the 1-st Block

C_j \ C_i	C_1	C_2	C_3	C_4
C_1	1.00	0.26	0.27	-0.12
C_2	0.32	1.00	0.38	-0.17
C_3	0.29	0.34	1.00	0.01
C_4	-0.15	-0.24	-0.01	1.00

FFN of the 1-st Block

C_j \ C_i	C_1	C_2	C_3	C_4
C_1	1.00	-0.31	0.41	-0.12
C_2	-0.29	1.00	-0.13	0.01
C_3	0.40	0.09	1.00	0.12
C_4	-0.09	0.02	0.11	1.00

Attention of the 4-th Block

C_j \ C_i	C_1	C_2	C_3	C_4
C_1	1.00	0.07	-0.66	-0.19
C_2	0.15	1.00	0.12	-0.13
C_3	-0.66	0.14	1.00	0.21
C_4	-0.21	-0.16	0.21	1.00

FFN of the 4-th Block

C_j \ C_i	C_1	C_2	C_3	C_4
C_1	1.00	0.22	0.27	-0.20
C_2	0.28	1.00	0.41	-0.40
C_3	0.37	0.38	1.00	-0.19
C_4	-0.18	-0.51	-0.30	1.00

Fig. 1. (a) Examples of PET images at different Centers. There are domain shifts between centers. (b) The interference metric $\mathcal{I}_{i,j}$ [16] of the center C_j on the center C_i at the 1-st/4-th blocks as examples. The red value indicates that C_j has a negative impact on C_i, and the green value indicates that C_j has a positive impact on C_i.

2 Method

2.1 Center Interference Issue

Due to the non-identical data distribution across centers, different centers with shared parameters may conflict with each other in the optimization process. To verify this hypothesis, we train a baseline Transformer with 15 base blocks (Fig. 2 (b)) over four centers. Following the paper [16], we calculate the gradient direction interference metric $\mathcal{I}_{i,j}$ of the j-th center C_j on the i-th center C_i. As shown in Fig. 1 (b), interference is observed between different centers at different layers. This will lead to inconsistent optimization and inevitably degrade the model performance. Details of $\mathcal{I}_{i,j}$ [16] are shown in the **supplement**.

2.2 Network Architecture

The overall architecture of our DRMC is shown in Fig. 2 (a). DRMC firstly applies a 3×3×3 convolutional layer for shallow feature extraction. Next, the shallow feature is fed into N blocks with dynamic routing (DRBs), which are expected to handle the interference between centers and adaptively extract the deep feature with high-frequency information. The deep feature then passes through another 3×3×3 convolutional layer for final image synthesis. In order to alleviate the burden of feature learning and stabilize training, DRMC adopts global residual learning as suggested in the paper [20] to estimate the image residual from different centers. In the subsequent subsection, we will expatiate the dynamic routing strategy as well as the design of the DRB.

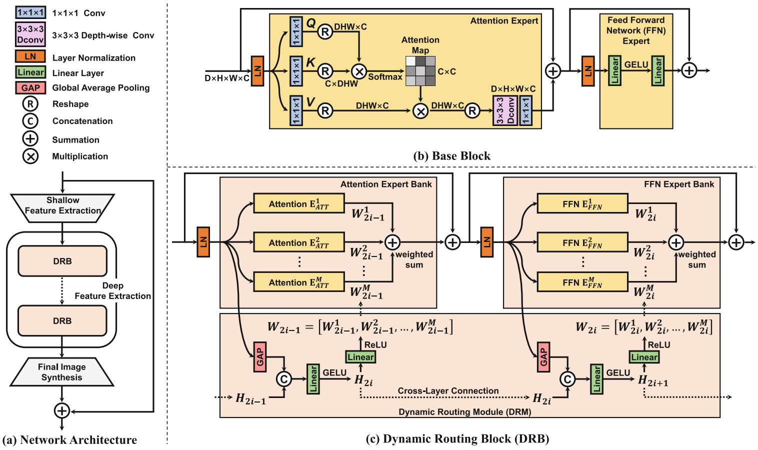


Fig. 2. The framework of our proposed DRMC

2.3 Dynamic Routing Strategy

We aim at alleviating the center interference issue in deep feature extraction. Inspired by prior generalist models [13,14,16], we specifically propose a novel dynamic routing strategy for multi-center PET image synthesis. The proposed dynamic routing strategy can be flexibly adapted to various network architectures, such as CNN and Transformer. To utilize the recent advance in capturing global contexts using Transformers [9], without loss of generality, we explore the application of the dynamic routing strategy to a Transformer block, termed dynamic routing block (DRB, see Fig. 2 (c)). We will introduce our dynamic routing strategy in detail from four parts: base expert foundation, expert number scaling, expert dynamic routing, and expert sparse fusion.

Base Expert Foundation. As shown in Fig. 2 (b), we first introduce an efficient base Transformer block (base block) consisting of an attention expert and a feed-forward network (FFN) expert. Both experts are for basic feature extraction and transformation. To reduce the complexity burden of the attention expert, we follow the paper [9] to perform global channel attention with linear complexity instead of spatial attention [21]. Notably, as the global channel attention may ignore the local spatial information, we introduce depth-wise convolutions to emphasize the local context after applying attention. As for the FFN expert, we make no modifications to it compared with the standard Transformer block [21]. It consists of a 2-layer MLP with GELU activation in between.

Expert Number Scaling. Center interference is observed on both attention experts and FFN experts at different layers (see Fig. 1 (b)). This indicates that a single expert can not be simply shared by all centers. Thus, we increase the number of experts in the base block to M to serve as expert candidates for different centers. Specifically, each Transformer block has an attention expert bank $\mathbf{E}_{ATT} = [\mathbf{E}^1_{ATT}, \mathbf{E}^2_{ATT}, ..., \mathbf{E}^M_{ATT}]$ and an FFN expert bank $\mathbf{E}_{FFN} = [\mathbf{E}^1_{FFN}, \mathbf{E}^2_{FFN}, ..., \mathbf{E}^M_{FFN}]$, both of which have M base experts. However, it does not mean that we prepare specific experts for each center. Although using center-specific experts can address the interference problem, it is hard for the model to exploit the shared knowledge across centers, and it is also difficult to generalize to new centers that did not emerge in the training stage [16]. To address this, we turn to different combinations of experts.

Expert Dynamic Routing. Given a bank of experts, we route data from different centers to different experts so as to avoid interference. Prior generalist models [13,14,16] in high-level vision tasks have introduced various routing strategies to weigh and select experts. Most of them are independently conditioned on the information of the current layer feature, failing to take into account the connectivity of neighboring layers. Nevertheless, PET image synthesis is a dense prediction task that requires a tight connection of adjacent layers for accurate voxel-wise intensity regression. To mitigate the potential discontinuity [13], we propose a dynamic routing module (DRM, see Fig. 2 (c)) that builds cross-layer connection for expert decisions. The mechanism can be formulated as:

$$W = \mathbf{ReLU}(\mathbf{MLP}([\mathbf{GAP}(X), H])), \tag{1}$$

where X denotes the input; $\mathbf{GAP}(\cdot)$ represents the global average pooling operation to aggregate global context information of the current layer; H is the hidden representation of the previous MLP layer. ReLU activation generates sparsity by setting the negative weight to zero. W is a sparse weight used to assign weights to different experts.

In short, DRM sparsely activates the model and selectively routes the input to different subsets of experts. This process maximizes collaboration and meanwhile mitigates the interference problem. On the one hand, the interference across centers can be alleviated by sparsely routing X to different experts (with positive weights). The combinations of selected experts can be thoroughly different across centers if violent conflicts appear. On the other hand, experts in the same bank still cooperate with each other, allowing the network to best utilize the shared knowledge across centers.

Expert Sparse Fusion. The final output is a weighted sum of each expert's knowledge using the sparse weight $W = [W^1, W^2, ..., W^M]$ generated by DRM. Given an input feature X, the output $\hat{X}$ of an expert bank can be obtained as:

$$\hat{X} = \sum_{m=1}^{M} W^m \cdot \mathbf{E}^m(X), \tag{2}$$

where $\mathbf{E}^m(\cdot)$ represents an operator of $\mathbf{E}^m_{ATT}(\cdot)$ or $\mathbf{E}^m_{FFN}(\cdot)$.

Table 1. Multi-Center PET Dataset Information

Center		Institution	Type	Lesion	System	Tracer	Dose	DRF	Spacing (mm^3)	Shape	Train	Test
C_{kn}	C_1	I_1	Whole Body	Yes	PolarStar m660	^{18}F-FDG	293MBq	12	3.15×3.15×1.87	192×192×*slices*	20	10
	C_2	I_2	Whole Body	Yes	PolarStar Flight	^{18}F-FDG	293MBq	4	3.12×3.12×1.75	192×192×*slices*	20	10
	C_3 [22]	I_3	Whole Body	Yes	United Imaging uEXPLORER	^{18}F-FDG	296MBq	10	1.67×1.67×2.89	256×256×*slices*	20	10
	C_4 [22]	I_4	Whole Body	Yes	Siemens Biograph Vision Quadra	^{18}F-FDG	296MBq	10	1.65×1.65×1.65	256×256×*slices*	20	10
C_{ukn}	C_5	I_5	Brain	No	PolarStar m660	^{18}F-FDG	293MBq	4	1.18×1.18×1.87	256×256×*slices*	–	10
	C_6	I_6	Whole Body	Yes	PolarStar m660	^{18}F-FDG	293MBq	12	3.15×3.15×1.87	192×192×*slices*	–	10

2.4 Loss Function

We utilize the Charbonnier loss [23] with hyper-parameter ϵ as 10^{-3} to penalize pixel-wise differences between the full-dose (Y) and estimated ($\hat{Y}$) PET images:

$$\mathcal{L} = \sqrt{\left\|Y - \hat{Y}\right\|^2 + \epsilon^2}. \tag{3}$$

3 Experiments and Results

3.1 Dataset and Evaluation

Full-dose PET images are collected from 6 different centers (C_1–C_6) at 6 different institutions[1]. The data of C_3 and C_4 [22] are borrowed from the Ultra-low Dose PET Imaging Challenge[2], while the data from other centers were privately collected. The key information of the whole dataset is shown in Table 1. Note that C_1–C_4 are for both training and testing. We denote them as C_{kn} as these centers are known to the generalist model. C_5 and C_6 are unknown centers (denote as C_{ukn}) that are only for testing the model generalizability. The low-dose PET data is generated by randomly selecting a portion of the raw scans based on

[1] I_1 and I_5 are Peking Union Medical College Hospital; I_2 is Beijing Hospital; I_3 is Department of Nuclear Medicine, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine; I_4 is Department of Nuclear Medicine, University of Bern; I_6 is Beijing Friendship Hospital.

[2] Challenge site: https://ultra-low-dose-pet.grand-challenge.org/. The investigators of the challenge contributed to the design and implementation of DATA, but did not participate in analysis or writing of this paper. A complete listing of investigators can be found at:https://ultra-low-dose-pet.grand-challenge.org/Description/.

the dose reduction factor (DRF), such as 25% when DRF=4. Then we reconstruct low-dose PET images using the standard OSEM method [24]. Since the voxel size differs across centers, we uniformly resample the images of different centers so that their voxel size becomes 2×2×2 mm^3. In the training phase, we unfold images into small patches (uniformly sampling 1024 patches from 20 patients per center) with a shape of 64×64×64. In the testing phase, the whole estimated PET image is acquired by merging patches together.

To evaluate the model performance, we choose the PSNR metric for image quantitative evaluation. For clinical evaluation, to address the accuracy of the standard uptake value (SUV) that most radiologists care about, we follow the paper [3] to calculate the bias of SUV_{mean} and SUV_{max} (denoted as B_{mean} and B_{max}, respectively) between low-dose and full-dose images in lesion regions.

Table 2. Results on C_{kn}. The **Best** and the Second-Best Results are Highlighted. *: Significant Difference at $p < 0.05$ between Comparison Method and Our Method.

Methods		PSNR↑					B_{mean}↓					B_{max}↓				
		C_1	C_2	C_3	C_4	Avg	C_1	C_2	C_3	C_4	Avg	C_1	C_2	C_3	C_4	Avg
(i)	3D-cGAN	47.30*	44.97*	45.15*	43.08*	45.13*	0.0968*	0.0832	0.0795*	0.1681*	0.1069*	0.1358*	0.1696*	0.1726*	0.2804*	0.1896*
	3D CVT-GAN	47.46*	45.17*	45.94*	44.04*	45.65*	0.0879	0.0972*	0.0594*	0.1413*	0.0965*	0.1178*	0.1591*	0.1652*	0.2224*	0.1661*
(ii)	FedAVG	47.43*	44.62*	45.61*	43.75*	45.35*	0.0985*	0.0996*	0.1006*	0.2202*	0.1122*	0.1459*	0.1546*	0.2011*	0.2663*	0.1920*
	FL-MRCM	47.81*	45.56*	46.10*	44.31*	45.95*	0.0939*	0.0929*	0.0631*	0.1344*	0.0961*	0.1571*	0.1607*	0.1307*	0.1518*	0.1501*
	FTL	48.05*	45.62*	46.01*	44.75*	46.11*	0.0892	0.0945*	0.0587*	0.0895	0.0830*	0.1243*	0.1588*	0.0893	0.1436	0.1290*
	DRMC	**49.48**	**46.32**	**46.71**	**45.01**	**46.88**	**0.0844**	**0.0792**	**0.0491**	**0.0880**	**0.0752**	**0.1037**	**0.1313**	**0.0837**	**0.1431**	**0.1155**

Table 3. Results on C_{ukn}.

Methods		PSNR↑		B_{mean}↓		B_{max}↓	
		C_5	C_6	C_5	C_6	C_5	C_6
(i)	3D-cGAN	26.53*	46.07*	–	0.1956*	–	0.1642*
	3D CVT-GAN	27.11*	46.03*	–	0.1828	–	0.1686*
(ii)	FedAVG	27.09*	46.48*	–	0.1943*	–	0.2291*
	FL-MRCM	25.38*	47.08*	–	0.1998*	–	0.1762*
	FTL	27.38*	48.05*	–	0.1898*	–	0.1556*
	DRMC	**28.54**	**48.26**	–	**0.1814**	–	**0.1483**

Table 4. Routing Ablation Results.

Methods	C_{kn}			C_{ukn}		
	PSNR↑	B_{mean}↓	B_{max}↓	PSNR↑	B_{mean}↓	B_{max}↓
w/o H	46.64*	0.0907*	0.1436*	38.23*	0.1826	0.1548*
Softmax	46.70*	0.0849*	0.1277*	38.33	0.1864*	0.1524*
Top-2 Gating	46.61*	0.0896*	0.1295*	38.38	0.1867*	0.1564*
DRMC	**46.88**	**0.0752**	**0.1155**	**38.40**	**0.1814**	**0.1483**

Table 5. Comparison results for Specialized Models and Generalist Models.

Methods		Train Centers	PNSR↑					B_{mean}↓					B_{max}↓				
			Test Centers				Avg.	Test Centers				Avg.	Test Centers				Avg.
			C_1	C_2	C_3	C_4		C_1	C_2	C_3	C_4		C_1	C_2	C_3	C_4	
Specialized Model	Baseline	C_1	48.89*	45.06*	43.94*	41.55*	44.86*	0.0849	0.0949*	0.1490*	0.2805*	0.1523*	0.1207*	0.1498*	0.3574*	0.4713*	0.2748*
		C_2	47.05*	46.08*	43.82*	41.53*	44.62*	0.0933*	**0.0557***	0.1915*	0.2247*	0.1413*	0.1326*	**0.1243***	0.3275*	0.4399*	0.2561*
		C_3	44.04*	41.00*	46.52*	44.07*	44.11*	0.2366*	0.2111*	**0.0446**	0.1364*	0.1572*	0.4351*	0.5567*	**0.0729***	0.1868*	0.3129*
		C_4	44.41*	41.39*	46.01*	44.95	44.29*	0.2462*	0.2063*	0.0897*	0.0966*	0.1597*	0.4887*	0.5882*	0.1222*	0.1562*	0.3388*
Generalist	Baseline	C_1, C_2, C_3, C_4	47.59*	44.73*	46.02*	44.20*	45.64*	0.0924*	0.0839*	0.0844*	0.1798*	0.1101*	0.1424*	0.1424*	0.1579*	0.2531*	0.1740*
Model	DRMC	C_1, C_2, C_3, C_4	**49.48**	**46.32**	**46.71**	**45.01**	**46.88**	**0.0844**	0.0792	0.0491	**0.0880**	**0.0752**	**0.1037**	0.1313	0.0837	**0.1431**	**0.1155**

3.2 Implementation

Unless specified otherwise, the intermediate channel number, expert number in a bank, and Transformer block number are 64, 3, and 5, respectively. We employ

Adam optimizer with a learning rate of 10^{-4}. We implement our method with Pytorch using a workstation with 4 NVIDIA A100 GPUs with 40GB memory (1 GPU per center). In each training iteration, each GPU independently samples data from a single center. After the loss calculation and the gradient back-propagation, the gradients of different GPUs are then synchronized. We train our model for 200 epochs in total as no significant improvement afterward.

3.3 Comparative Experiments

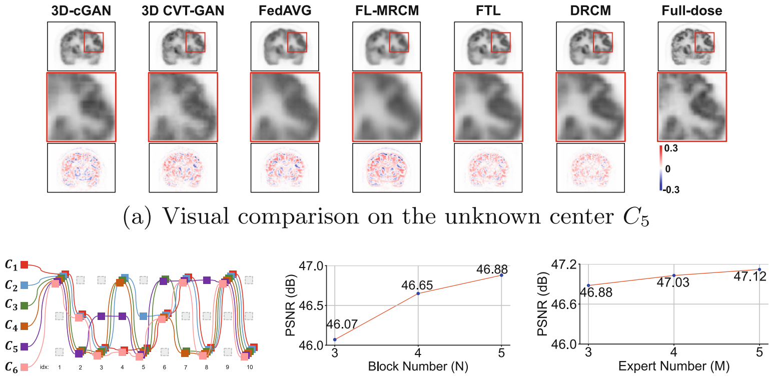
We compare our method with five methods of two types. (i) 3D-cGAN [1] and 3D CVT-GAN [10] are two state-of-the-art methods for single center PET image synthesis. (ii) FedAVG [11,12], FL-MRCM [11], and FTL [7] are three federated learning methods for privacy-preserving multi-center medical image synthesis. All methods are trained using data from C_{kn} and tested over both C_{kn} and C_{ukn}. For methods in (i), we regard C_{kn} as a single center and mix all data together for training. For federated learning methods in (ii), we follow the "**Mix**" mode (upper bound of FL-based methods) in the paper [11] to remove the privacy constraint and keep the problem setting consistent with our multi-center study.

Comparison Results for Known Centers. As can be seen in Table 2, in comparison with the second-best results, DRMC boosts the performance by 0.77 dB PSNR, 0.0078 B_{mean}, and 0.0135 B_{max}. This is because our DRMC not only leverages shared knowledge by sharing some experts but also preserves center-specific information with the help of the sparse routing strategy. Further evaluation can be found in the **supplement**.

Comparison Results for Unknown Centers. We also test the model generalization ability to unknown centers C_5 and C_6. C_5 consists of normal brain data (without lesion) that is challenging for generalization. As the brain region only occupies a small portion of the whole-body data in the training dataset but has more sophisticated structure information. C_6 is a similar center to C_1 but has different working locations and imaging preferences. The quantitative results are shown in Table 3 and the visual results are shown in Fig. 1 (a). DRMC achieves the best results by dynamically utilizing existing experts' knowledge for generalization. On the contrary, most comparison methods process data in a static pattern and unavoidably produce mishandling of out-of-distribution data. Furthermore, we investigate model's robustness to various DRF data, and the results are available in the **supplement**.

3.4 Ablation Study

Specialized Model vs. Generalist Model. As can be seen in Table 5, the baseline model (using 15 base blocks) individually trained for each center acquires good performance on its source center. But it suffers performance drop on other centers. The baseline model trained over multiple centers greatly enhances the overall results. But due to the center interference issue, its performance on a specific center is still far from the corresponding specialized model.



(a) Visual comparison on the unknown center C_5

(b) Top-1 Expert (c) PSNR/N (d) PSNR/M

Fig. 3. Figures of different experiments.

DRMC mitigates the interference with dynamic routing and achieves comparable performance to the specialized model of each center.

Ablation Study of Routing Strategy. To investigate the roles of major components in our routing strategy, we conduct ablation studies through (i) removing the condition of hidden representation H that builds cross-layer connection, and replacing ReLU activation with (ii) softmax activation [14] and (iii) top-2 gating [13]. The results are shown in Table 4. We also analyze the interpretability of the routing by showing the distribution of different layers' top-1 weighted experts using the testing data. As shown in Fig. 3 (b), different centers show similarities and differences in the expert distribution. For example, C_6 shows the same distribution with C_1 as their data show many similarities, while C_5 presents a very unique way since brain data differs a lot from whole-body data.

Ablation Study of Hyperparameters. In Fig. 3 (c) and (d), we show ablation results on expert number (M) and block number (N). We set M=3 and N=5, as it has realized good performance with acceptable computational complexity.

4 Conclusion

In this paper, we innovatively propose a generalist model with dynamic routing (DRMC) for multi-center PET image synthesis. To address the center interference issue, DRMC sparsely routes data from different centers to different experts. Experiments show that DRMC achieves excellent generalizability.

References

1. Wang, Y., et al.: 3d conditional generative adversarial networks for high-quality pet image estimation at low dose. NeuroImage **174**, 550–562 (2018)
2. Xiang, L., et al.: Deep auto-context convolutional neural networks for standard-dose pet image estimation from low-dose pet/MRI. Neurocomputing **267**, 406–416 (2017)

3. Zhou, L., Schaefferkoetter, J., Tham, I., Huang, G., Yan, J.: Supervised learning with cyclegan for low-dose FDG pet image denoising. Med. Image Anal. **65**, 101770 (2020)
4. Zhou, Y., Yang, Z., Zhang, H., Chang, E.I.C., Fan, Y., Xu, Y.: 3d segmentation guided style-based generative adversarial networks for pet synthesis. IEEE Trans. Med. Imaging **41**(8), 2092–2104 (2022)
5. Luo, Y., Zhou, L., Zhan, B., Fei, Y., Zhou, J., Wang, Y.: Adaptive rectification based adversarial network with spectrum constraint for high-quality pet image synthesis. Med. Image Anal. **77**, 102335 (2021)
6. Chaudhari, A., et al.: Low-count whole-body pet with deep learning in a multicenter and externally validated study. NPJ Digit. Med. **4**, 127 (2021)
7. Zhou, B., et al.: Federated transfer learning for low-dose pet denoising: a pilot study with simulated heterogeneous data. IEEE Trans. Radiat. Plasma Med. Sci. **7**(3), 284–295 (2022)
8. Luo, Y., et al.: 3D transformer-GAN for high-quality PET reconstruction. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12906, pp. 276–285. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87231-1_27
9. Jang, S.I., et al.: Spach transformer: spatial and channel-wise transformer based on local and global self-attentions for pet image denoising, September 2022
10. Zeng, P., et al.: 3D CVT-GAN: a 3d convolutional vision transformer-GAN for PET reconstruction, pp. 516–526, September 2022
11. Guo, P., Wang, P., Zhou, J., Jiang, S., Patel, V.M.: Multi-institutional collaborations for improving deep learning-based magnetic resonance image reconstruction using federated learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2423–2432, June 2021
12. McMahan, H.B., Moore, E., Ramage, D., Hampson, S., et al.: Communication-efficient learning of deep networks from decentralized data. arXiv preprint arXiv:1602.05629 (2016)
13. Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q., Hinton, G., Dean, J.: Outrageously large neural networks: the sparsely-gated mixture-of-experts layer, January 2017
14. Wang, X., Cai, Z., Gao, D., Vasconcelos, N.: Towards universal object detection by domain attention. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7289–7298 (2019)
15. Zhu, X., et al.: Uni-perceiver: pre-training unified architecture for generic perception for zero-shot and few-shot tasks. arXiv preprint arXiv:2112.01522 (2021)
16. Zhu, J., et al.: Uni-perceiver-MOE: learning sparse generalist models with conditional MOEs. In: Oh, A.H., Agarwal, A., Belgrave, D., Cho, K. (eds.) Advances in Neural Information Processing Systems (2022)
17. Wang, P., et al.: OFA: unifying architectures, tasks, and modalities through a simple sequence-to-sequence learning framework. CoRR abs/2202.03052 (2022)
18. Yu, T., Kumar, S., Gupta, A., Levine, S., Hausman, K., Finn, C.: Gradient surgery for multi-task learning. arXiv preprint arXiv:2001.06782 (2020)
19. Han, Y., Huang, G., Song, S., Yang, L., Wang, H., Wang, Y.: Dynamic neural networks: a survey, February 2021
20. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a gaussian denoiser: residual learning of deep CNN for image denoising. IEEE Trans. Image Process. **26**(7), 3142–3155 (2017)
21. Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems, pp. 5998–6008 (2017)

22. Xue, S., et al.: A cross-scanner and cross-tracer deep learning method for the recovery of standard-dose imaging quality from low-dose pet. Eur. J. Nucl. Med. Mol. Imaging **49**, 1619–7089 (2022)
23. Charbonnier, P., Blanc-Feraud, L., Aubert, G., Barlaud, M.: Two deterministic half-quadratic regularization algorithms for computed imaging. In: Proceedings of 1st International Conference on Image Processing. vol. 2, pp. 168–172 (1994)
24. Hudson, H., Larkin, R.: Accelerated image reconstruction using ordered subsets of projection data. IEEE Trans. Med. Imaging **13**(4), 601–609 (1994)