Parameter Efficient Fine Tuning for Multi-scanner PET to PET Reconstruction

Yumin Kim* Gayoon Choi* Seong Jae Hwang†

Department of Artificial Intelligence, Yonsei University, Seoul, Republic of Korea
{yumin, gynchoi17, seongjae}@yonsei.ac.kr

Abstract. Reducing scan time in Positron Emission Tomography (PET) imaging while maintaining high-quality images is crucial for minimizing patient discomfort and radiation exposure. Due to the limited size of datasets and distribution discrepancy across scanners in medical imaging, fine-tuning in a parameter-efficient and effective manner is on the rise. Motivated by the potential of Parameter Efficient Fine-Tuning (PEFT), we aim to address these issues by effectively leveraging PEFT to improve limited data and GPU resource issues in multi-scanner setups. In this paper, we introduce **PETITE**, **P**arameter **E**fficient Fine-**T**uning for Mult**I**-scanner PE**T** to PET R**E**construction, which represents the optimal PEFT combination when independently applying encoder-decoder components to each model architecture. To the best of our knowledge, this study is the first to systematically explore the efficacy of diverse PEFT techniques in medical imaging reconstruction tasks via prevalent encoder-decoder models. This investigation, in particular, brings intriguing insights into PETITE as we show further improvements by treating the encoder and decoder separately and mixing different PEFT methods, namely, Mix-PEFT. Using multi-scanner PET datasets comprised of five different scanners, we extensively test the cross-scanner PET scan time reduction performances (i.e., a model pre-trained on one scanner is fine-tuned on a different scanner) of 21 feasible Mix-PEFT combinations to derive optimal PETITE. We show that training with less than 1% parameters using PETITE performs on par with full fine-tuning (i.e., 100% parameter). Code is available at: https://github.com/MICV-yonsei/PETITE

Keywords: Parameter Efficient Fine-Tuning · Vision Transformer · Positron Emission Tomography (PET) · PET reconstruction

1 Introduction

Positron Emission Tomography (PET) is an in vivo nuclear medicine technique using radiotracers for early diagnosis of Alzheimer's and Parkinson's diseases [2,4]. Despite their clinical value, long-time scans can cause motion artifacts that lead to discomfort for the patients [11]. To address this issue, PET reconstruction

* Equal contribution † Correponding author

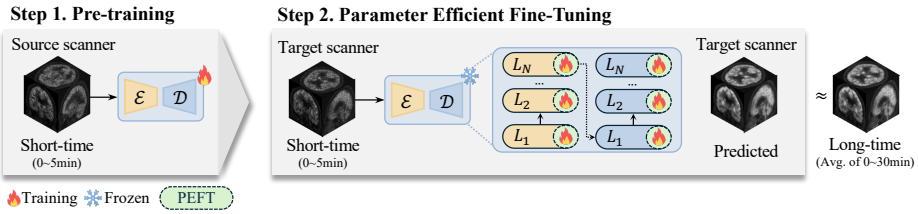

Fig. 1. The overview of ***PETITE***: Scheme for single source-target settings in PET scan time reduction with PEFT.

emerges as a remarkable technique, enhancing image quality significantly without extending scanning time [3, 20, 26]. It has continuously evolved through various deep learning methods, enabling the reconstruction of high-quality scans from short scans that closely match those acquired from longer scans.

An array of generative models to achieve PET reconstruction have been developed. For instance, Wang *et al.* [23] leveraged a Generative Adversarial Networks (GAN) [8] to synthesize images that are hard to discriminate with high-dose PET. Recently, the rise of Vision Transformer (ViT) [5] has also demonstrated its potential for PET reconstruction. Expanding on this integration, Luo *et al.* [20] inserts ViTs in CNN Encoder-Decoder to take advantage of both CNN and Transformer and Zeng *et al.* [26] adopt convolution in the self-attention mechanism to reduce semantic ambiguity. However, these models encounter efficiency challenges due to the process known as full fine-tuning (Full-FT), which involves updating all layers of a large model, resulting in increased training time and higher GPU resource consumption.

In medical imaging, the diversity of equipment and protocols poses persistent practical challenges, significantly impacting data generalization and model application. This issue manifests similarly in the PET domain due to discrepancies in scanner manufacturers, imaging facilities, or protocol types that complicate application from hospital to hospital. Such variations limit the generalization of models trained on a particular dataset to others, requiring fine-tuning in hospitals with limited datasets. Specifically, among fine-tuning methods, generalization in multi-scanner is also an important issue.

Consequently, due to the scarcity of datasets and inefficient large-scale models, some works progressively adopted *Parameter-Efficient Fine-Tuning* (PEFT). PEFT approaches significantly decrease storage requirements and computational costs by freezing most parameters of a pre-trained model and selectively fine-tuning a limited set of parameters. Despite considerable research on PEFT [7, 16, 19], there are few attempts at medical imaging, with prior studies primarily focusing on classification tasks [6, 10].

Recognizing the gaps in applying PEFT within medical imaging, particularly in reconstruction tasks, we pioneer the use of the PEFT methodology for PET reconstruction from short-time scans, aiming to reduce scan duration and enhance reconstruction quality [1, 15]. Specifically in scenarios with rela-

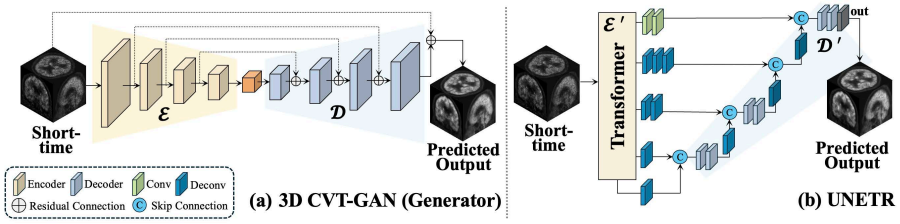

Fig. 2. The pipeline of the encoder-decoder structure of each ViT-based model. (a) 3D CVT-GAN [26] consists of a generator with a ViT-based encoder and decoder. Only the first three layers of the encoder and the first two layers of the decoder are trained. (b) UNETR [9] consists of a ViT-based encoder and a CNN-based decoder.

tively limited data availability, we aim to enhance reconstruction performance across diverse scanners using various PEFT methods. Consequently, we introduce **PETITE**, **P**arameter-**E**fficient Fine-**T**uning for Mult**I**-scanner PE**T** to PET R**E**construction, which represents the optimal PEFT combination when applying the encoder-decoder approach to each model architecture. PETITE uses fewer than 1% of the parameters for PET reconstruction using short-time scans aimed at reducing scan time, as detailed in Fig. 1, which describes its scheme for single source-target settings. Furthermore, we investigate the optimal experimental settings for reconstruction models through various PEFT approaches. Specifically, we explore the synergistic effects of applying diverse PEFT methods independently to the distinct encoder and decoder components, the process which is defined as **Mix-PEFT**.

Contributions. Our main contributions are as follows: **(1)** We leverage the PEFT methodology in a medical reconstruction task to reduce the scan time of PET images on scanners with different dimensions, voxel spacing, and institutions. To the best of our knowledge, this extensive study represents the first application of the PEFT methodology within the field of medical imaging. **(2)** Upon experimenting with possible Mix-PEFT, we found that using less than 1% of parameters can achieve performance comparable to Full-FT, carefully considering encoder and decoder architecture. **(3)** We provide novel insights into the optimal PEFT settings tailored for the reconstruction model.

2 Methodology

We briefly describe relevant PET reconstruction models and outline parameter efficient fine-tuning (PEFT) approaches (Sec. 2.1). Then, we detail Mix-PEFT to consider encoder and decoder separately (Sec. 2.2).

Scan-time Reduction Model. (a) 3D CVT-GAN [26] aims to effectively integrate CNN and ViT technologies for high-quality PET reconstruction. This architecture replaces projection in multi-head attention from the linear to the

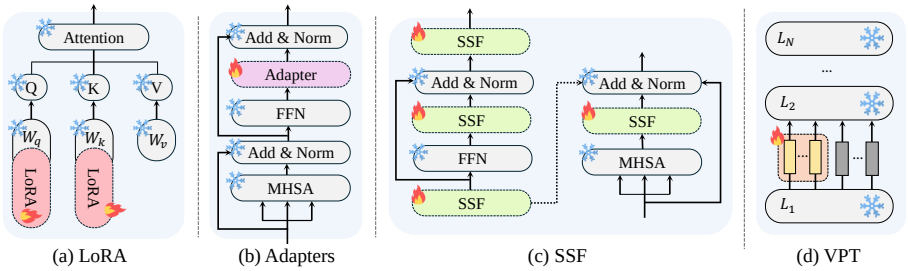

Fig. 3. Illustrations of the modified structures of PEFT methods.

convolutional [24], building encoder-decoder structure and combining the conditional GAN [21]. The architecture includes an encoder block for feature extraction and decoder blocks for restoring high-quality PET scans, capturing both local spatial features and global contexts from various network layers. During the pre-training step, the discriminator is also trained, but it is frozen in the PEFT step. (b) UNETR [9] comprises a ViT-based encoder for extracting representations that are merged with the CNN-based decoder through skip connections at multiple resolutions to predict outputs. For detailed structures into the positions and structures of the encoder and decoder in the model used, see Fig. 2.

2.1 PEFT for PET Scan-time Reduction

We describe the PEFT concept and describe the model-specific adjustments and their rationale. We have divided the PEFT category into two variants: Selective and Additive methods. The optimal hyperparameter values for PEFT are given in the supplementary.

Selective methods. This method includes (I) LayerNorm tuning, and (II) BitFit, and leverages the model's pre-training procedure without making major changes. This approach precisely fine-tunes specific layers or a segment of the original pre-trained model, without adding new modules, optimizing performance efficiently.

(I)*LayerNorm tuning* tunes parameters $\theta' \in \theta$ in the normalization layer for adapting intermediate statistics to align with target distributions.

(II)*BitFit* proposes only updates bias-term of the network. Based on our experimental results, BitFit tuning has been identified as one of the competitive baselines for PEFT. Building on this discovery, we found that combining BitFit tuning with all PEFT methodologies yields better performance than using original PEFT alone. Specifically, integrating BitFit with particularly effective techniques in our study, such as LoRA, SSF, and VPT described below, led to notable performance enhancements, as shown in Table 2(a). These findings indicate that BitFit and PEFT methods are effectively complementary.

Additive methods. The additive method can be categorized into four types [17]: (I) LoRA, (II) Adapters, (III) SSF, and (IV) VPT. This approach involves augmenting the pre-trained model with additional trainable layers, where only the parameters of these new layers are subject to training.

(I)*LoRA* injects learnable rank decomposition matrices into pre-trained weight matrices, under the hypothesis that, during adaptation, weight updates have a low intrinsic rank [13]. The low-rank decomposition matrix is scaled with a factor α that is constant to a row-rank r, and the number of trainable parameters depends on r. As shown in Fig. 3(a), we apply LoRA to the query and key matrices, since tuning only these matrices yields better results and a parameter reduction of 0.22% for the 3D CVT-GAN model and 0.03% for UNETR, resulting in better performance, with more simplicity than adjusting the query, key, and value matrices together. In UNETR, tuning with a learning rate higher than 1e-2 leads to a fall in local optima and hinders training. Analysis of rank revealed that $r = 1, 4, 8$ are stable in performance, whereas $r = 16$ resulted in decreased performance.

(II)*Adapters* introduces lightweight modules, adding fully connected networks after attention, and feed-forward layers in Transformer [12]. An adapter module typically comprises a linear down-projection, succeeded by a nonlinear activation function, and concluded with a linear up-projection, all integrated with a residual connection. The bottleneck architecture enables parameter reduction through a reduction factor (rf). As shown in Fig. 3(b), we add the Adapters only after the feed-forward layer to further reduce the computational cost according to Pfieffer [22]. We test adapters sizes in {1,4,8,16,32}. Analysis of rank revealed that r = 4,8 are stable in performance.

(III)*Scaling & Shifting Your Features (SSF)* modulates deep features x from the pre-trained model via linear transformations [18]. SSF module consists of the scale factor γ, dot product with x, and the shift factor β, added to x. It is injected after the multi-head attention layer, the feed-forward layer, and the normalization layer of the Transformer. As shown in Fig. 3(c), it is injected after every MLP, MHSA, and Layernorm module, scaling and shifting features from them during training, and can be re-parameterized at inference since it is a linear structure.

(IV)*Visual-Prompt Tuning (VPT)* introduces a small set of p continuous vectors in the embedding space of every encoder, tailored to learn task-specific information via attention. [14]. Since VPT-deep, which inserts prompt tokens into every encoder layer, does not align well with reconstruction tasks, our approach solely utilizes VPT-shallow. Additionally, we observed that inserting prompt tokens from the second encoder, bypassing the first, enhanced performance as depicted in Fig. 3(d). Inserting prompt tokens into the first encoder layer can disrupt representation learning, as it may interfere with focusing on blocks containing task-relevant information due to the variance in locations across pre-trained ViT models [25]. In 3D CVT-GAN, we inserted 8 and 32 tokens into the encoder, while in UNETR, we added 50 tokens. The function of VPT is as follows:

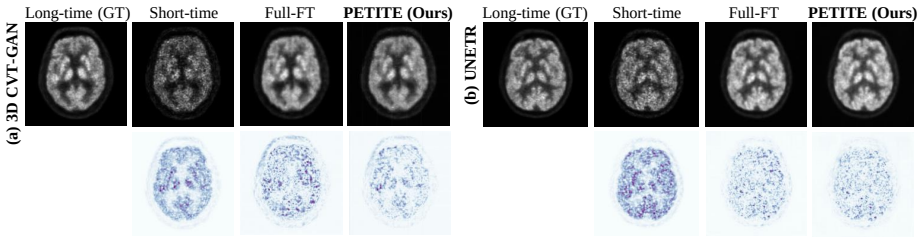

Fig. 4. Scan time reduction examples. *First row*: PET scans. *Second row*: error maps comparing the reconstructed PET scans to the ground-truth (GT).

$$[\mathbf{Z}_2, \mathbf{E}_2] = L_2([\mathbf{P}, \mathbf{E}_1]), \quad [\mathbf{Z}_i, \mathbf{E}_i] = L_i([\mathbf{Z}_{i-1}, \mathbf{E}_{i-1}]), \quad i = 3, 4, ..., N. \tag{1}$$

2.2 PETITE: Optimal Effectiveness of Mix-PEFT

Mix-PEFT refers to the approach of applying various PEFT methods independently to the encoder and decoder within each model, aiming to achieve and analyze the ideal synergistic effect. This method particularly highlights PEFT techniques that consistently provide positive impacts on both the encoder and decoder in two ViT-based models, the 3D CVT-GAN and UNETR models.

(I) ViT-based encoder-decoder: The 3D CVT-GAN applies VPT [14] in the encoder and LoRA [13] in the decoder, representing the optimal Mix-PEFT combination. By applying task-specific prompt tokens to the encoder via VPT, the 3D CVT-GAN is fine-tuned for specific tasks to improve feature extraction. Using VPT and LoRA together creates a synergistic effect, as both operate within the critical attention component of ViT. For parameter efficiency, the decoder tunes only the query and key matrices in the attention layer to produce high-quality PET scans.

(II) ViT-based encoder and a CNN-based decoder: The UNETR applies LoRA to the encoder and SSF [18] to the decoder, representing the optimal Mix-PEFT combination. SSF is best for CNNs. Tuning SSF in ViT with other PEFT methods results in a performance drop, as it hinders ViT's complex attention mechanism. While PEFT was originally designed for ViT layers, it has been adapted to be compatible with CNN-based decoders in our work.

The representations extracted from the encoder are matched to the target dataset's distribution by applying SSF to the CNN-based decoder, utilizing the features of the pre-trained model.

3 Experiments and Results

Datasets and Image Preprocessing. In our study, we used the Alzheimer's Disease Neuroimaging Initiative (ADNI) public dataset for PET imaging across

Table 1. Quantitative comparison of PSNR, SSIM, and NRMSE between PETITE (Ours) and other PEFT methods. *Best*: **Bold**; *Second best*: Underline

(a) 3D CVT-GAN [26]					(b) UNETR [9]				
Method	**% Param**	**PSNR**(↑)	**SSIM**(↑)	**NRMSE**(↓)	**Method**	**% Param**	**PSNR**(↑)	**SSIM**(↑)	**NRMSE**(↓)
No-FT	-	30.329	0.905	0.0327	No-FT	-	30.034	0.787	0.0388
Full-FT	100%	**31.410**	0.908	**0.0292**	Full-FT	100%	**31.746**	**0.884**	**0.0285**
LayerNorm	0.041%	31.304	**0.914**	**0.0292**	LayerNorm	0.037%	30.753	0.811	0.0308
BitFit	0.12%	31.291	0.910	0.0294	BitFit	0.07%	30.969	0.810	0.0299
LoRA	0.45%	30.982	0.909	0.0308	LoRA	0.46%	31.010	0.825	0.0299
Adapters	0.71%	31.108	0.913	0.0304	Adapters	0.44%	30.657	0.798	0.0311
SSF	0.16%	31.018	0.809	0.0298	SSF	0.11%	30.989	0.809	0.0298
VPT	0.02%	30.398	0.892	0.0331	VPT	0.038%	30.494	0.796	0.0318
PETITE *(Ours)*	0.32%	31.373	0.912	0.0298	**PETITE *(Ours)***	0.51%	31.696	0.865	0.0292

five different scanners and institutions with details on multi-scanner information in Supplementary. ^{18}F-Fluorodeoxyglucose (^{18}F-FDG) was injected at a dose of 185 MBq (5 mCi) for the scans, with each ADNI PET scan consisting of a sequence of six 5-minute frame scans (i.e., 0-5, ..., 25-30 minutes). The short-time scans PET is the first 5-minute scan (0-5) only. The long-time scans PET (GT) is generated by simply averaging these six five-minute frames. Our dataset was split into 30 training and 15 validation samples for pre-training, and 10 training and 15 validation samples for parameter efficient fine-tuning (PEFT). We processed the images using the MONAI Library, applying a random size crop to $64 \times 64 \times 64$ and normalizing the intensities of all reconstructed PET images to a 0-1 range.

Performance Assesment. If scanner 1 is used as the source, pre-training is performed on scanner 1, and PEFT on the other four scanners. This process is repeated for each scanner as the source, PEFT the others. Each source scanner is thus fine-tuned four times, resulting in 20 possible PEFT results across five scanners. After averaging the results for the source-target pairs, excluding the same scanner as the source, we averaged the five resulting values again. Additionally, we performed a 3-fold cross-validation, leading to further averaging across three iterations for robust results.

Evaluation Metrics. The peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), and normalized root mean squared error (NRMSE) are used as quantitative evaluation metrics. Among them, PSNR indicates estimation accuracy in terms of the logarithmic decibel scale, while SSIM and NRMSE represent the structural similarity and voxel-wise intensity differences between the ground-truth and predicted images, respectively.

Implementation Details. The hyperparameters we tuned include the number of epochs, batch size, learning rate, learning rate decay, learning rate scheduler, the rank value of LoRA [13], the reduction factor of Adapters, and the number of VPT prompt tokens [14]. All models were trained with a batch size of 6, using PyTorch and MONAI for implementation. Both models performed pre-training for 1000 epochs. The PEFT involved training the 3D CVT-GAN for 150 epochs and the UNETR for 200 epochs. The final performance was determined based on the epoch with the highest PSNR value. See supplementary for detailed hyperparameters for PEFT.

Table 2. Optimal Mix-PEFT (Ours) outperforms each PEFT method with or without BitFit. *Best*: **Bold**; *Second best*: Underline

(a) 3D CVT-GAN [26]						(b) UNETR [9]					
Method	**BitFit**	**% Param**	**PSNR**(↑)	**SSIM**(↑)	**NRMSE**(↓)	**Method**	**BitFit**	**% Param**	**PSNR**(↑)	**SSIM**(↑)	**NRMSE**(↓)
No-FT	-	-	30.329	0.905	0.0327	No-FT	-	-	30.034	0.787	0.0338
Full-FT	-	100%	**31.410**	0.908	**0.0292**	Full-FT	-	100%	**31.746**	**0.884**	**0.0282**
BitFit	✗	0.12%	31.291	0.910	0.0294	BitFit	✗	0.07%	30.969	0.810	0.0299
VPT	✗	0.02%	30.398	0.892	0.0331	LoRA (r=4)	✗	0.44%	31.010	0.825	0.0299
LoRA (r=8)	✗	0.72%	31.050	0.900	0.0300	SSF	✗	0.2%	30.897	0.808	0.0303
VPT	✓	0.14%	31.302	0.909	0.0300	LoRA (r=4)	✓	0.51%	31.010	0.825	0.0299
LoRA (r=8)	✓	0.61%	31.073	0.910	0.0307	SSF	✓	0.2%	30.897	0.808	0.0303
PETITE *(Ours)*	✗	0.20%	31.305	0.910	0.0307	PETITE *(Ours)*	✗	0.44%	31.193	0.819	0.0310
PETITE *(Ours)*	✓	0.32%	31.373	**0.912**	0.0298	**PETITE *(Ours)***	✓	0.51%	31.696	0.865	0.0292

3.1 Evaluation Results

Quantitative Experiments. We evaluated our proposed method on two ViT-based models: 3D CVT-GAN [26] and UNETR [9], employing various PEFT methods such as (1) LayerNorm tuning, (2) BitFit, (3) LoRA, (4) Adapters, (5) SSF, (6) VPT, and (7) PETITE (Ours). Specifically, when compared to one of the competitive baselines, Adapters, in the 3D CVT-GAN (as shown in Table 1(a)), our proposed PETITE approach achieved improvements of 0.263 in PSNR and 0.0006 in NRMSE, while utilizing 0.4% fewer parameters. Similarly, in the UNETR (as shown in Table 1(b)), PETITE demonstrated its efficacy as a parameter-efficient approach by improving the metrics PSNR, SSIM, and NRMSE by 1.039, 0.067, and 0.0019, respectively, while maintaining a parameter size comparable to that of Adapters.

Qualitative Experiments. In Fig. 4 the lighter color of the error map indicates a smaller error. As observed, the quantitative experiments of the 3D CVT-GAN model show that only the PSNR is comparable to the performance of Full-FT, Fig. 4, the error is smaller and closer to the long-time scans PET (GT) compared to Full-FT.

Ablation Study. To assess the impact of key components in our proposed PETITE, we carry out ablation studies on two models by considering the following configurations: (1) Baseline, (2) PEFT, (3) PEFT + BitFit, and (4) PETITE (Ours). Commonly, the (1) Baseline encompasses No-FT and Full-FT, while (2) original PEFT method is tuned on all layers, and the combination with (3) BitFit is explored to demonstrate the efficacy of tuning the original PEFT alongside BitFit. Although various combinations were possible, (4) Ours represents the best-performing combination identified through ablation studies for each model. As shown in Table 2, combining the PEFT method tailored for the encoder-decoder structure with BitFit tuning across all layers yields superior performance. Notably, within the UNETR model in Table 2(b), our approach, despite using 81.9% fewer parameters than the ViT-based encoder, demonstrates higher performance in terms of PSNR and SSIM by 0.169 and 0.07, respectively, thereby validating the effectiveness of the PETITE methodology.

4 Conclusions

Our comprehensive experiments on multi-scanner PET datasets have affirmed the effectiveness of PETITE, demonstrating that the synergetic effects of Mix-PEFT enable achieving results akin to full fine-tuning with less than 1% of parameters. This efficient approach not only mitigates issues arising from limited datasets and discrepancies in scanner distributions but also addresses the critical need to reduce PET scan times while maintaining image quality. The insights gained from separately addressing encoder and decoder components and integrating various PEFT methods highlight the potential of PETITE to innovate medical imaging reconstruction tasks. This study lays the groundwork for future research in parameter-efficient methodologies, foreseeing extensive exploration of parameter-efficient fine-tuning methods in the medical imaging field.

Acknowledgments. This work was supported in part by the IITP 2020-0-01361 (AI Graduate School Program at Yonsei University), NRF RS-2023-00262002, and NRF RS-2023-00219019 funded by Korean Government (MSIT).

Disclosure of Interests. The authors have no competing interests.

References

1. Anwar, S.M., Majid, M., Qayyum, A., Awais, M., Alnowami, M., Khan, M.K.: Medical image analysis using convolutional neural networks: a review. Journal of medical systems **42**, 1–13 (2018)
2. Becker, G., Müller, A., Braune, S., Büttner, T., Benecke, R., Greulich, W., Klein, W., Mark, G., Rieke, J., Thümler, R.: Early diagnosis of parkinson's disease. Journal of neurology **249**(Suppl 3), iii40–iii48 (2002)
3. Conti, M.: Focus on time-of-flight pet: the benefits of improved time resolution. European journal of nuclear medicine and molecular imaging **38**(6), 1147–1157 (2011)
4. Cummings, J.: The national institute on aging—alzheimer's association framework on alzheimer's disease: application to clinical trials. Alzheimer's & Dementia **15**(1), 172–178 (2019)
5. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale (2020)
6. Dutt, R., Ericsson, L., Sanchez, P., Tsaftaris, S.A., Hospedales, T.M.: Parameter-efficient fine-tuning for medical image analysis: The missed opportunity. CoRR **abs/2305.08252** (2023), https://doi.org/10.48550/arXiv.2305.08252
7. Edalati, A., Tahaei, M.S., Kobyzev, I., Nia, V.P., Clark, J.J., Rezagholizadeh, M.: Krona: Parameter efficient tuning with kronecker adapter. CoRR **abs/2212.10650** (2022), https://doi.org/10.48550/arXiv.2212.10650
8. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. Advances in neural information processing systems **27** (2014)
9. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R., Xu, D.: Unetr: Transformers for 3d medical image segmentation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 574–584 (2022)

10. He, J., Zhou, C., Ma, X., Berg-Kirkpatrick, T., Neubig, G.: Towards a unified view of parameter-efficient transfer learning. CoRR **abs/2110.04366** (2021), https://arxiv.org/abs/2110.04366
11. Herzog, H., Tellmann, L., Fulton, R., Stangier, I., Kops, E.R., Bente, K., Boy, C., Hurlemann, R., Pietrzyk, U.: Motion artifact reduction on parametric pet images of neuroreceptor binding. Journal of Nuclear Medicine **46**(6), 1059–1065 (2005)
12. Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., Attariyan, M., Gelly, S.: Parameter-efficient transfer learning for nlp. In: International Conference on Machine Learning. pp. 2790–2799. PMLR (2019)
13. Hu, E.J., yelong shen, Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: LoRA: Low-rank adaptation of large language models. In: International Conference on Learning Representations (2022), https://openreview.net/forum?id=nZeVKeeFYf9
14. Jia, M., Tang, L., Chen, B.C., Cardie, C., Belongie, S., Hariharan, B., Lim, S.N.: Visual prompt tuning. In: European Conference on Computer Vision. pp. 709–727. Springer (2022)
15. Li, X., Zhang, L., Wu, Z., Liu, Z., Zhao, L., Yuan, Y., Liu, J., Li, G., Zhu, D., Yan, P., et al.: Artificial general intelligence for medical imaging. arXiv preprint arXiv:2306.05480 (2023)
16. Li, X.L., Liang, P.: Prefix-tuning: Optimizing continuous prompts for generation pp. 4582–4597 (2021)
17. Lialin, V., Deshpande, V., Rumshisky, A.: Scaling down to scale up: A guide to parameter-efficient fine-tuning. CoRR **abs/2303.15647** (2023), https://doi.org/10.48550/arXiv.2303.15647
18. Lian, D., Zhou, D., Feng, J., Wang, X.: Scaling & shifting your features: A new baseline for efficient model tuning. Advances in Neural Information Processing Systems **35**, 109–123 (2022)
19. Liu, X., Ji, K., Fu, Y., Du, Z., Yang, Z., Tang, J.: P-tuning v2: Prompt tuning can be comparable to fine-tuning universally across scales and tasks. CoRR **abs/2110.07602** (2021), https://arxiv.org/abs/2110.07602
20. Luo, Y., Wang, Y., Zu, C., Zhan, B., Wu, X., Zhou, J., Shen, D., Zhou, L.: 3d transformer-gan for high-quality pet reconstruction. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24. pp. 276–285. Springer (2021)
21. Mirza, M., Osindero, S.: Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784 (2014)
22. Pfeiffer, J., Kamath, A., Rücklé, A., Cho, K., Gurevych, I.: Adapterfusion: Non-destructive task composition for transfer learning. CoRR **abs/2005.00247** (2020), https://arxiv.org/abs/2005.00247
23. Wang, Y., Yu, B., Wang, L., Zu, C., Lalush, D.S., Lin, W., Wu, X., Zhou, J., Shen, D., Zhou, L.: 3d conditional generative adversarial networks for high-quality pet image estimation at low dose. Neuroimage **174**, 550–562 (2018)
24. Wu, H., Xiao, B., Codella, N., Liu, M., Dai, X., Yuan, L., Zhang, L.: Cvt: Introducing convolutions to vision transformers. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 22–31 (2021)
25. Yoo, S., Kim, E., Jung, D., Lee, J., Yoon, S.: Improving visual prompt tuning for self-supervised vision transformers. In: International Conference on Machine Learning. pp. 40075–40092. PMLR (2023)

26. Zeng, P., Zhou, L., Zu, C., Zeng, X., Jiao, Z., Wu, X., Zhou, J., Shen, D., Wang, Y.: 3d cvt-gan: A 3d convolutional vision transformer-gan for pet reconstruction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 516–526. Springer (2022)