Frequency Strikes Back: Boosting Parameter-Efficient Foundation Model Adaptation for Medical Imaging

Son T. Ly[1] and Hien V. Nguyen[1]

Department of Electronics and Computer Engineering, University of Houston, USA
stly@cougarnet.uh.edu

Abstract. Adapting vision transformer (ViT) foundation models with parameter-efficient fine-tuning (PEFT) has become increasingly popular in medical imaging, enabling efficient adaptation while updating only a small subset of parameters. However, existing PEFT methods process tokens independently, overlooking cross-token dependencies and limiting their ability to capture global contextual information. To address these challenges, we propose FreqFiT, a novel Frequency-based Fine-Tuning module inserted between ViT blocks to enhance model adaptability. FreqFiT is effective and seamlessly integrates with existing PEFT methods to improve their performance. We evaluate FreqFiT across 2D and 3D medical imaging datasets, such as PAPILA, HAM10000, ADNI-1.5T, and COVID-CT-MD. It improves accuracy 9% and AUC 10%, surpassing the original PEFT methods on both MedMAE and DINOv2 backbones. Despite using only $\leq 1.2\%$ of full fine-tuning parameters, FreqFiT achieves state-of-the-art medical imaging adaptation efficiently. The source code is available here.

Keywords: Foundation models · PEFTs · FFT

1 Introduction

Deep learning has transformed medical image analysis, but training models from scratch is challenging due to privacy constraints, dataset imbalances, and the high cost of expert annotations [22, 4]. Foundation models (FMs), which leverage large-scale pretraining followed by fine-tuning on domain-specific data, have proven effective across various biomedical imaging modalities like ultrasound, histopathology, and radiology [7, 16, 20, 14].

However, full fine-tuning of FMs requires significant computational resources and large annotated datasets, which are often unavailable in medical domains. To address these limitations, parameter-efficient fine-tuning (PEFT) methods, which update only a fraction of model parameters, have gained attention for their efficiency in domain adaptation tasks with minimal computational cost [6, 8, 3]. PEFT techniques have achieved comparable performance to full fine-tuning, updating less than 1% of parameters, making them ideal for scenarios where full fine-tuning is not feasible due to privacy concerns or limited resources [10, 4].

Despite their efficiency, PEFT methods process tokens independently, ignoring crucial cross-token dependencies and limiting their ability to capture global context. This limitation weakens their adaptability in medical imaging, where understanding complex relationships across tokens is essential. Addressing these challenges is key to enhancing model robustness and ensuring reliable, generalizable performance across diverse medical scenarios.

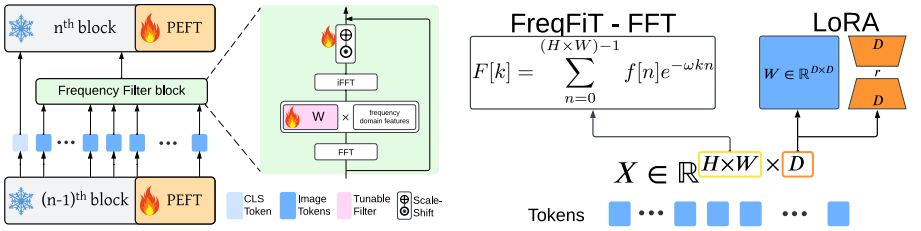
To overcome this challenge, we propose FreqFiT, a frequency-based fine-tuning module that enhances adaptability while remaining computationally efficient. Integrated between ViT blocks, FreqFiT aggregates statistics across all the tokens by leveraging the frequency domain to capture subtle patterns that spatial-domain methods often miss. Its plug-and-play design ensures compatibility with existing PEFT techniques like LoRA, delivering consistent performance improvements across diverse medical imaging tasks without requiring architectural changes, making it suitable for both research and clinical applications. We evaluate FreqFiT across 2D and 3D medical imaging datasets. It surpasses the original PEFT methods on both medical and natural imaging pre-trained backbones, despite using only $\leq 1.2\%$ of full fine-tuning parameters. Our main contributions are as follows:

1. We propose FreqFit, a simple and effective frequency-based fine-tuning module that seamlessly integrates with existing PEFT methods to enhance model adaptation.
2. We provide theoretical support for why FreqFit can capture image features that existing PEFT methods cannot, such as information across all the tokens.
3. We provide empirical results on 2D and 3D medical imaging datasets on both medical and natural imaging pre-trained foundation models.

2 Related Works

Parameter-Efficient Fine-Tuning (PEFT). PEFT methods reduce the cost of adapting large-scale vision models. PEFT methods, such as LoRA [6], AdaLoRA [24], BOFT [13], address this challenge by fine-tuning a small subset of model parameters. LoRA introduces low-rank decomposition to reduce the parameter overhead, while AdaLoRA dynamically adjusts the rank during training for improved efficiency. BOFT focuses on block-oriented tuning to preserve spatial hierarchies. Despite their effectiveness, these methods often overlook the frequency-domain properties of medical images, potentially limiting their ability to capture subtle anatomical variations. However, these spatial-domain methods may overlook the information across tokens, restricting their ability to model long-range interactions and capture subtle anatomical variations essential for medical imaging.

PEFT in Medical Imaging. Fully fine-tuning foundation models for medical imaging is challenging due to modality variations and limited labeled data. Recent works have explored PEFT to address these issues. [4] benchmarked 17 PEFT methods, showing up to 22% gains in low-data scenarios. [12] applied LoRA to chest radiography, outperforming full fine-tuning in 13/18 tasks with $<1\%$



(a) Overview of FreqFiT (b) Difference between FreqFiT and LoRA

Fig. 1. (a) Overview of FreqFiT. (b) Methodology differences between FreqFiT, utilizing Fast Fourier Transform (FFT), and LoRA. Given input tokens $X \in \mathbb{R}^{H \times W \times D}$, while LoRA operates on the channel dimension D in the spatial domain, FreqFiT, using FFT, processes both the $H \times W$ and D in the frequency domain. This highlights how FreqFiT and PEFT methods like LoRA complement each other. See Sec. 3 for details.

tunable parameters. [9] used PEFT methods such as LoRA to explore fairness in medical applications. Other studies have applied PEFT to specialized tasks, such as volumetric organ segmentation [19], multi-scanner PET reconstruction [10], and organ segmentation across imaging modalities [2]. These studies underscore PEFT's effectiveness for medical imaging adaptation while minimizing computational overhead. However, these studies do not consider the fall short of PEFT methods mentioned above.

FreqFiT distinguishes itself from prior approaches by uniquely integrating PEFT with frequency-based adaptations, enabling robust medical imaging adaptation. Its $O(1)$ parameter complexity ensures scalability to both 2D and 3D inputs, all while preserving the integrity of the foundation model. This makes FreqFiT particularly well-suited for scenarios where retraining the entire model is not feasible.

3 FreqFit - Frequency Fine-tuning

Here, we formally introduce the frequency tuning method, called FreqFiT, as illustrated in Fig. 1. FreqFiT integrates a frequency operator consisting of a filter basis followed by a residual connection, between ViT blocks. Given an input token $X \in \mathbb{R}^{H \times W \times D}$, we perform FFT along the spatial dimensions to transform the input into X_c, as in Eq. 1. We modulate X_c by multiplying it with the learnable filter $K \in \mathbb{R}^{H \times W \times D}$, which has the same dimensions as X_c, as in Eq. 2. Finally, we convert the modulated spectral features $\tilde{X}$ back to the spatial domain using the inverse FFT, as shown in Eq. 3, and add a residual connection from the original input X. This process can be mathematically presented as follows:

$$X_c = \mathcal{F}(X) \in \mathbb{C}^{H \times W \times D} \tag{1}$$

$$\widetilde{X_c} = K \odot X_c \tag{2}$$

$$\tilde{X} \leftarrow \mathcal{F}^{-1}(\widetilde{X_c}) \tag{3}$$

$$\hat{X} = X + \alpha \odot \tilde{X} + \beta \tag{4}$$

where, $\mathcal{F}$ and $\mathcal{F}^{-1}$ are the fast Fourier transform (FFT) and its inverse. $K \in \mathbb{R}^{H\times W\times D}$ denotes a learnable filter, α and β are the scale and shift factors, and $\odot$ is the Hadamard product.

How Does FreqFiT Improve Performance? The basic idea behind frequency-based tuning FreqFiT is to learn the interactions among spatial locations in the frequency domain. We provide the theoretical foundation that underscores the importance of incorporating FreqFiT into the existing fine-tuning paradigm. See also Fig. 1 for better understanding.

Proposition 1. *FreqFiT with $O(1)$ parameters can create a feature transformation that spatial-domain parameter-efficient fine-tuning methods cannot replicate.*

Proof. Here, we employ LoRA to ease the proof as this can be generalized to other PEFT methods. Given the input tokens $X \in \mathbb{R}^{H\times W\times D}$, where each token is a D-dimensional vector spread across an $H \times W$ spatial grid. The LoRA transformation and FreqFiT are as:

$$X_{LoRA} = W_0X + BAX \tag{5}$$

$$X_{FreqFiT} = X + \mathcal{F}^{-1}(\mathcal{F}(X) \odot K) \tag{6}$$

where, $K \in \mathbb{C}^{H\times W\times D}$, and $B \in \mathbb{R}^{D\times r}$ and $A \in \mathbb{R}^{r\times D}, r \ll D$ are the low-rank matrices. For simplicity, we do not present α and β parameters for FreqFiT.

For FreqFiT to replicate LoRA transformation, and vice versa, there must exist a filter K such that:

$$\mathcal{F}^{-1}(\mathcal{F}(X) \odot K) = BAX \tag{7}$$

However, this is not generally possible because of the following reasons: *(i)* Filter K and BA operate in different domains, i.e., frequency and spatial domains, respectively. *(ii)* K is a 3D filter that modulates information in both the tokens 2-dimensional $H \times W$ and the channel dimension D. *(iii)* BA, in Eq. 5 is a token-specific modification, where the D-dimensional representation of each token is updated. This modification captures relationships within and across channels, such as correlations or dependencies among the features in D.

As a result, FreqFiT introduces implicit cross-token interaction via aggregated statistics across all the tokens. Whereas, LoRA operates locally, emphasizing token-specific updates in the channel dimension.

While LoRA transformation is not full rank, FreqFiT transformation is full rank. The Fourier transform of $X \in \mathbb{R}^{H\times W\times D}$ is given by:

$$\hat{X} = U_H X U_W^T \qquad \textit{// Fourier transform} \tag{8}$$

$$\hat{Y} = F\hat{X} = F\left(U_H X U_W^T\right) \qquad \textit{// Multiply with } F \tag{9}$$

$$Y = U_H^H \hat{Y} (U_W^H)^T \qquad \textit{// Inverse Fourier transform} \tag{10}$$

where $U_H \in C^{H\times H}$ and $U_W \in C^{W\times W}$ are the unitary Fourier transform matrix of row and column grid. The U_H^H and U_W^H are the conjugate transpose of U_H and U_W, respectively, i.e., $U_H^H U_H = I$ and $U_W^H U_W = I$. Thus, if the input X is

full-rank and F is a full-rank diagonal matrix (no zero entries for all frequency components), since U is unitary, the rank of X is preserved in the frequency domain, then the result X must be full-rank.

FreqFit has $O(1)$ parameter complexity, as the frequency modulation filter F can be parameterized efficiently, focusing only on essential frequency components, regardless the input dimensions of X.

Proposition 2. *Combining FreqFiT with spatial-domain PEFT methods can create a feature transformation that cannot be achieved by FreqFiT or any spatial-domain PEFT method alone.*

Proof. Let $f_{comb}(X) = f_{\text{FreqFiT}}(f_{\text{LoRA}}(X))$, where f_{FreqFiT} and f_{LoRA}, as defined in Proposition 1. Assume, for contradiction, that $f \in f_{\text{LoRA}} \cup f_{\text{FreqFiT}}$. If $f_{comb} \in f_{\text{LoRA}}$, then

$$f_{comb}(X) = W_0 X + B'A'X \tag{11}$$

for some low-rank A', B'. However, this form cannot capture the non-local, frequency-aware masking introduced by f_{FreqFiT}, which operates globally across spatial tokens—beyond the representational scope of spatial PEFT. On the other hand, if $f_{comb} \in f_{\text{FreqFiT}}$, then

$$f_{comb}(X) = X + \mathcal{F}^{-1}(\mathcal{F}(X) \odot K') \tag{12}$$

for some $K' \in \mathbb{C}^{H \times W \times D}$. Yet, this form cannot represent the intermediate channel-wise transformation ABX applied before the FFT, as the Fourier transform is linear. Hence, no $f_{comb} \in f_{\text{FreqFiT}}$ or f_{LoRA} can reproduce the composite transformation f_{LoRA} and f_{LoRA}, and therefore:

$$f_{comb} \notin f_{\text{LoRA}} \cup f_{\text{FreqFiT}}. \tag{13}$$

Proposition 2 provides a compelling rationale for combining two complementary approaches: FreqFiT and traditional PEFT methods like LoRA. By leveraging their distinct strengths, this combination enables a significantly more effective fine-tuning strategy. Our experimental results across diverse datasets strongly validate this theory, demonstrating substantial improvements over using either method alone.

4 Experiments

We evaluate our method using foundation models pre-trained on natural or medical image datasets, encompassing diverse learning paradigms, including DINOv2 [17] for self-supervised learning on natural images and MedMAE [23] for masked autoencoding pretraining on medical images.

To assess FreqFiT's versatility, we integrate it with various state-of-the-art PEFT methods—LoRA [6], AdaLoRA [24], FourierFT [5], and BOFT [13]. These methods serve as baselines to evaluate how FreqFiT enhances parameter-efficient adaptation across different pre-training strategies and datasets. All experiments

Table 1. Comparison between the state-of-the-arts PEFT methods and their FreqFiT-enhanced counterpart. The gray-shaded **Bold** indicates that the performances in both ACC and AUC are equal to or better than the original method.

	Tuned Params		MedMAE								DINOv2							
	M	% to Full	PAPILA		HAM10000		ADNI-1.5T		COVID-CT-MD		PAPILA		HAM10000		ADNI-1.5T		COVID-CT-MD	
			ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC
LoRA	0.44	0.5%	78.6	0.78	87.7	0.88	76.0	0.81	73.5	0.60	91.1	0.93	86.1	0.89	75.0	0.83	93.5	0.97
FreqFit-LoRA	0.88	1.0%	**84.0**	**0.88**	**87.1**	**0.88**	**81.7**	**0.85**	**78.2**	**0.82**	**94.6**	**0.97**	**86.8**	**0.91**	**81.7**	**0.85**	**94.1**	**0.98**
AdaLoRA	0.66	0.8%	59.0	0.64	84.2	0.86	49.0	0.61	78.2	0.82	83.9	0.91	82.9	0.88	75.0	0.78	85.3	0.91
FreqFit-AdaLoRA	1.11	1.2%	**76.8**	**0.80**	**85.5**	**0.88**	**78.8**	**0.77**	**78.8**	**0.82**	**89.3**	**0.92**	**88.1**	**0.90**	**75.0**	**0.82**	**87.6**	**0.92**
BOFT	0.11	0.13%	80.4	0.81	82.2	0.87	72.1	0.76	77.6	0.84	89.3	0.93	85.5	0.88	73.1	0.79	93.5	0.97
FreqFit-BOFT	0.55	0.6%	**87.5**	**0.84**	**83.9**	**0.87**	**80.9**	**0.80**	**80.0**	**0.83**	**93.0**	**0.96**	**88.7**	**0.91**	**80.8**	**0.86**	**97.6**	**0.99**
FourierFT	0.025	0.03%	76.8	0.70	84.6	0.84	60.6	0.70	75.3	0.69	89.3	0.93	85.5	0.88	74.0	0.80	73.5	0.70
FreqFit-FourierFT	0.47	0.5%	**85.1**	**0.82**	84.2	0.87	**78.8**	**0.85**	**79.4**	**0.83**	**93.0**	**0.96**	**88.2**	**0.91**	75.0	0.78	**89.0**	**0.94**

use default hyperparameters, including low-rank configurations, from Hugging Face's PEFT library [15], applied to all linear layers.

For downstream tasks, we conduct binary classification experiments on diverse 2D and 3D medical imaging datasets: PAPILA [11] for papillary thyroid carcinoma, HAM10000 [21] for dermatoscopic lesion analysis, COVID-CT-MD [1] for COVID-19 diagnosis from CT scans, and ADNI-1.5T [18] for Alzheimer's detection using MRI. Data splits and preprocessing follow [9, 4].

5 Main Resutls

Table 1 presents fine-tuning results on various imaging datasets, showing that integrating FreqFiT-enhanced methods consistently improves accuracy (ACC) and area under the curve (AUC) across various tasks and foundation models. FreqFiT-enhanced methods outperform their baselines across all datasets, demonstrating its effectiveness in handling the complexities of medical imaging data.

For the MedMAE [23], FreqFiT-BOFT achieves top performance on PAPILA (ACC: 87.5%, AUC: 0.84) and COVID-CT-MD (ACC: 80.0%, AUC: 0.83), surpassing the original BOFT by 7.1% and 2.4% in accuracy, respectively. FreqFiT-LoRA also improves on ADNI-1.5T, raising accuracy from 76.0% to 81.7% and AUC from 0.81 to 0.85, demonstrating its ability to capture subtle clinical features.

For the DINOv2 [17], FreqFiT methods show strong improvements across tasks. FreqFiT-LoRA achieves 94.6% accuracy and 0.97 AUC on PAPILA, surpassing LoRA by 3.5% and 0.04, respectively. FreqFiT-BOFT reaches the highest performance on COVID-CT-MD (ACC: 97.6%, AUC: 0.99), highlighting the benefits of frequency-domain adaptation in volumetric data. Compared to stronger baselines like FourierFT, FreqFiT variants consistently outperform, with FreqFiT-FourierFT gaining 5.8% on PAPILA and 5.9% on COVID-CT-MD.

Although FreqFiT introduces additional parameters compared to original PEFT methods, the total parameter count remains small relative to full tuning. FreqFiT-based methods tune less than 1.2% of the full tuning ($\sim 85M$), showing that FreqFiT remains parameter-efficient while enhancing performance.

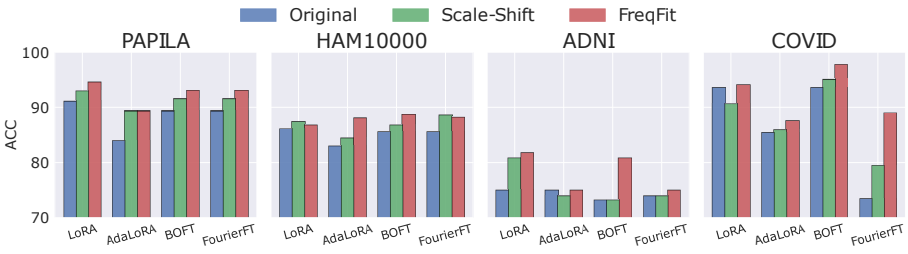


Fig. 2. FreqFit and Scale-Shift-based methods, a simple linear transformation, consistently perform better than the original PEFT methods.

Overall, these results highlight FreqFiT's robustness and versatility across different PEFT strategies, backbone models, and imaging modalities, achieving consistent gains with minimal computational overhead.

6 Ablations

Frequency-based Tuning is better than Linear Transformation. We compare FreqFiT with Scaling-Shifting-enhanced and original PEFT methods on medical imaging datasets. Scaling-Shifting applies a linear transformation with learnable scaling and shifting factors between ViT blocks, adjusting the amplitude and mean of input features. As shown in Fig. 2, FreqFiT outperforms Scaling-Shifting, achieving an average gain of 1.3% in accuracy (ACC) and 0.04 in AUC across all methods. This improvement highlights FreqFit's superior ability to align input features with the frozen model parameters, especially when dealing with the diverse anatomical structures and imaging modalities inherent in medical datasets. Unlike Scaling-Shifting, which adjusts features globally, FreqFit's frequency-based approach captures a broad spectrum of spatial dependencies by modulating both high- and low-frequency components. This enables the model to better represent long-range structural patterns and fine-grained pathological details—crucial for tasks like tumor boundary delineation in PAPILA or subtle lesion detection in HAM10000.

While Scaling-Shifting approach offers computational efficiency and effectively addresses moderate distribution shifts, its global feature adjustment limits its capacity to capture intricate spatial dependencies and frequency variations. As observed in COVID-CT-MD and ADNI-1.5T, SS-based methods underperform compared to their FreqFit-enhanced counterparts, particularly in scenarios requiring the capture of subtle structural abnormalities. While Scaling-Shifting provides modest gains in PEFT, especially in domains with minimal domain gaps, it lacks the nuanced adaptability that FreqFit offers. The consistent improvements with FreqFit, even in challenging 3D medical imaging tasks, underscore the benefits of frequency-domain transformations for robust and adaptable parameter-efficient fine-tuning.

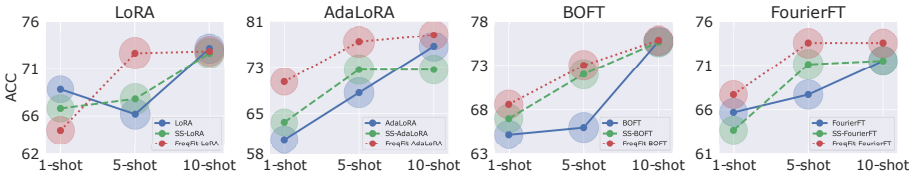


Fig. 3. Comparison FreqFiT, Scale-Shift-based, and PEFT methods on few-shot settings. The size of regions around markers are proportional to the corresponding AUC scores.

It's also important to note that while Scaling-Shifting still lags behind FreqFiT, it improves performance over original PEFT methods, highlighting that existing PEFT techniques could overlook features' distribution alignment between frozen ViT blocks, leading to suboptimal feature alignment.

Few-shot Setting. In Fig. 3, the results on the HAM10000 dataset highlight the effectiveness of FreqFit-enhanced PEFT methods across different few-shot learning settings. In the 1-shot scenario, FreqFit consistently outperforms both original PEFT and scale-shift PEFT (SS-PEFT) variants. Notably, FreqFit-LoRA achieves the highest accuracy (73.1%) and a strong AUC (0.63), improving upon both LoRA (68.8%, 0.57 AUC) and SS-LoRA (66.8%, 0.59 AUC). Similarly, FreqFit-AdaLoRA and FreqFit-BOFT demonstrate superior performance compared to their respective baselines, indicating that frequency modulation enhances adaptation even in extreme data scarcity.

In the 5-shot setting, FreqFit continues to show significant improvements. FreqFit-AdaLoRA achieves the best overall performance with 77.4% accuracy and 0.78 AUC, surpassing AdaLoRA (68.6%, 0.72 AUC) and SS-AdaLoRA (72.6%, 0.71 AUC). Likewise, FreqFit-LoRA (72.6%, 0.79 AUC) and FreqFit-BOFT (73.0%, 0.71 AUC) outperform their respective baselines, demonstrating the robustness of frequency-based adaptations.

In the 10-shot setting, FreqFit methods maintain competitive performance. While the gap between methods narrows as more labeled data becomes available, FreqFit-AdaLoRA still achieves the highest accuracy (78.6%) among all approaches. Meanwhile, FreqFit-LoRA (72.8%) and FreqFit-BOFT (75.8%) sustain their improvements over standard PEFTs. The results suggest that FreqFit enhances parameter-efficient fine-tuning by preserving essential information across tokens while mitigating the instability caused by frozen model blocks.

Overall, these results validate FreqFit as an effective enhancement for PEFT in medical imaging, particularly in low-data regimes where traditional PEFT methods struggle with token interactions.

7 Limitations

Despite its effectiveness, FreqFiT requires slightly more parameters than standard PEFT methods. While it only uses $\sim 1\%$ fraction of full fine-tuning parameters, the additional overhead may limit its deployment in resource-restricted settings.

Future work could explore further parameter efficiency optimizations to balance performance gains with computational cost.

8 Conclusion

In this study, we introduce FreqFiT, a novel frequency-tuning module that enhances PEFT methods by modifying ViT frequencies, addressing overlooked tokens information. Easily integrated into existing fine-tuning approaches, FreqFiT operates in the frequency domain, enabling models to capture subtle patterns more effectively. Extensive experiments across 2D and 3D medical imaging tasks demonstrate its effectiveness, consistently improving performances. Despite minimal parameter overhead, FreqFiT achieves state-of-the-art adaptation, surpassing standard PEFT methods in medical imaging.

Acknowledgments. This research was supported by the National Institutes of Health (NIH) under Award Number 5R01CA277739-03. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Parnian Afshar, Shahin Heidarian, Nastaran Enshaei, Farnoosh Naderkhani, Moezedin Javad Rafiee, Anastasia Oikonomou, Faranak Babaki Fard, Kaveh Samimi, Konstantinos N Plataniotis, and Arash Mohammadi. Covid-ct-md, covid-19 computed tomography scan dataset applicable in machine learning and deep learning. *Scientific Data*, 8(1):121, 2021. 6
2. Naif Alkhunaizi, Faris Almalik, Rouqaiah Al-Refai, Muzammal Naseer, and Karthik Nandakumar. Probing the efficacy of federated parameter-efficient fine-tuning of vision transformers for medical image classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 236–245. Springer, 2024. 3
3. Shoufa Chen, Chongjian Ge, Zhan Tong, Jiangliu Wang, Yibing Song, Jue Wang, and Ping Luo. Adaptformer: Adapting vision transformers for scalable visual recognition. *arXiv:2205.13535*, 2022. 1
4. Raman Dutt, Linus Ericsson, Pedro Sanchez, Sotirios A Tsaftaris, and Timothy Hospedales. Parameter-efficient fine-tuning for medical image analysis: The missed opportunity. *arXiv preprint arXiv:2305.08252*, 2023. 1, 2, 6
5. Ziqi Gao, Qichao Wang, Aochuan Chen, Zijing Liu, Bingzhe Wu, Liang Chen, and Jia Li. Parameter-efficient fine-tuning with discrete fourier transform. *arXiv preprint arXiv:2405.03003*, 2024. 5
6. Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. *arXiv:2106.09685*, 2021. 1, 2, 5

7. Joana Palés Huix, Adithya Raju Ganeshan, Johan Fredin Haslum, Magnus Söderberg, Christos Matsoukas, and Kevin Smith. Are natural domain foundation models useful for medical image classification? In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 7634–7643, 2024. 1
8. Menglin Jia, Luming Tang, Bor-Chun Chen, Claire Cardie, Serge Belongie, Bharath Hariharan, and Ser-Nam Lim. Visual prompt tuning. pages 709–727, 2022. 1
9. Ruinan Jin, Zikang Xu, Yuan Zhong, Qiongsong Yao, Qi Dou, S Kevin Zhou, and Xiaoxiao Li. Fairmedfm: fairness benchmarking for medical imaging foundation models. *arXiv preprint arXiv:2407.00983*, 2024. 3, 6
10. Yumin Kim, Gayoon Choi, and Seong Jae Hwang. Parameter efficient fine tuning for multi-scanner pet to pet reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 518–528. Springer, 2024. 1, 3
11. Oleksandr Kovalyk, Juan Morales-Sánchez, Rafael Verdú-Monedero, Inmaculada Sellés-Navarro, Ana Palazón-Cabanes, and José-Luis Sancho-Gómez. Papila: Dataset with fundus images and clinical data of both eyes of the same patient for glaucoma assessment. *Scientific Data*, 9(1):291, 2022. 6
12. Chenyu Lian, Hong-Yu Zhou, Yizhou Yu, and Liansheng Wang. Less could be better: Parameter-efficient fine-tuning advances medical vision foundation models. *arXiv preprint arXiv:2401.12215*, 2024. 2
13. Weiyang Liu, Zeju Qiu, Yao Feng, Yuliang Xiu, Yuxuan Xue, Longhui Yu, Haiwen Feng, Zhen Liu, Juyeon Heo, Songyou Peng, et al. Parameter-efficient orthogonal finetuning via butterfly factorization. *arXiv preprint arXiv:2311.06243*, 2023. 2, 5
14. Jun Ma and Bo Wang. Towards foundation models of biological image segmentation. *Nature Methods*, 20(7):953–955, 2023. 1
15. Sourab Mangrulkar, Sylvain Gugger, Lysandre Debut, Younes Belkada, Sayak Paul, and Benjamin Bossan. Peft: State-of-the-art parameter-efficient fine-tuning methods. https://github.com/huggingface/peft, 2022. 6
16. Michael Moor, Oishi Banerjee, Zahra Shakeri Hossein Abad, Harlan M Krumholz, Jure Leskovec, Eric J Topol, and Pranav Rajpurkar. Foundation models for generalist medical artificial intelligence. *Nature*, 616(7956):259–265, 2023. 1
17. Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023. 5, 6
18. Ronald Carl Petersen, Paul S Aisen, Laurel A Beckett, Michael C Donohue, Anthony Collins Gamst, Danielle J Harvey, CR Jack Jr, William J Jagust, Leslie M Shaw, Arthur W Toga, et al. Alzheimer's disease neuroimaging initiative (adni) clinical characterization. *Neurology*, 74(3):201–209, 2010. 6
19. Julio Silva-Rodríguez, Jose Dolz, and Ismail Ben Ayed. Towards foundation models and few-shot parameter-efficient fine-tuning for volumetric organ segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 213–224. Springer, 2023. 3
20. Hari Sowrirajan, Jingbo Yang, Andrew Y Ng, and Pranav Rajpurkar. Moco pretraining improves representation and transferability of chest x-ray models. In *Medical Imaging with Deep Learning*, pages 728–744. PMLR, 2021. 1
21. Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific data*, 5(1):1–9, 2018. 6

22. Martin J Willemink, Wojciech A Koszek, Cailin Hardell, Jie Wu, Dominik Fleischmann, Hugh Harvey, Les R Folio, Ronald M Summers, Daniel L Rubin, and Matthew P Lungren. Preparing medical imaging data for machine learning. *Radiology*, 295(1):4–15, 2020. 1
23. Junfei Xiao, Yutong Bai, Alan Yuille, and Zongwei Zhou. Delving into masked autoencoders for multi-label thorax disease classification. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3588–3600, 2023. 5, 6
24. Qingru Zhang, Minshuo Chen, Alexander Bukharin, Nikos Karampatziakis, Pengcheng He, Yu Cheng, Weizhu Chen, and Tuo Zhao. Adalora: Adaptive budget allocation for parameter-efficient fine-tuning. *arXiv preprint arXiv:2303.10512*, 2023. 2, 5