Regularized Low-Rank Adaptation for Few-Shot Organ Segmentation

Ghassen Baklouti[1,2,✉], Julio Silva-Rodríguez[1], Jose Dolz[1,2], Houda Bahig[2], and Ismail Ben Ayed[1,2]

[1]ÉTS Montréal
✉ghassen.baklouti.1@ens.etsmtl.ca
[2]Centre de Recherche du Centre Hospitalier de l'Université de Montréal (CRCHUM)

Abstract. Parameter-efficient fine-tuning (PEFT) of pre-trained foundation models is increasingly attracting interest in medical imaging due to its effectiveness and computational efficiency. Among these methods, Low-Rank Adaptation (LoRA) is a notable approach based on the assumption that the adaptation inherently occurs in a low-dimensional subspace. While it has shown good performance, its implementation requires a fixed and unalterable rank, which might be challenging to select given the unique complexities and requirements of each medical imaging downstream task. Inspired by advancements in natural image processing, we introduce a novel approach for medical image segmentation that dynamically adjusts the intrinsic rank during adaptation. Viewing the low-rank representation of the trainable weight matrices as a singular value decomposition, we introduce an l_1 sparsity regularizer to the loss function, and tackle it with a proximal optimizer. The regularizer could be viewed as a penalty on the decomposition rank. Hence, its minimization enables to find task-adapted ranks automatically. Our method is evaluated in a realistic few-shot fine-tuning setting, where we compare it first to the standard LoRA and then to several other PEFT methods across two distinguishable tasks: **base organs** and **novel organs**. Our extensive experiments demonstrate the significant performance improvements driven by our method, highlighting its efficiency and robustness against suboptimal rank initialization. Our code is publicly available: https://github.com/ghassenbaklouti/ARENA.

Keywords: Few-shot learning · Segmentation · Low-rank adaptation

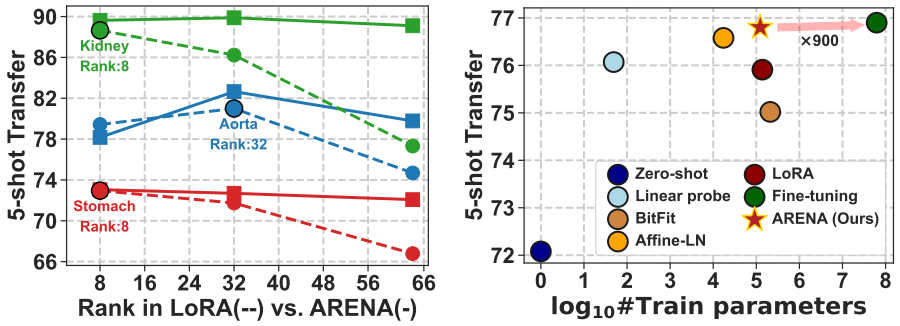
1 Introduction

Despite the remarkable success of deep neural networks for automatic volumetric organ segmentation [19], these have shown limited flexibility. Concretely, they are usually specialized in specific tasks, requiring large annotated datasets for training and heavy computational demands. As a result, medical image segmentation has been hindered by the necessity of assembling large, curated datasets for deployment [7], which is particularly expensive due to the manual labor effort

required for annotating volumetric data [31]. However, a paradigm shift is underway, led by *foundation models*, and more particularly medical-specialized pre-trained models [26,18,30,21]. These models are pre-trained on large-scale datasets that leverage heterogeneous, multi-domain samples and tackle multiple tasks within specific medical domains. Thanks to such intensive pre-training, they have demonstrated remarkable generalization properties [25,27,18], as well as flexible adaptation to downstream tasks, while requiring minimal supervision [10,27]. These properties highlight foundation models as a promising path toward a more efficient deployment of automatic segmentation solutions [32,24,37], with recent open-access models developed for volumetric organ segmentation [20,28,27,18], which are pre-trained on an assembly of annotated datasets segmenting multiple base structures. Nevertheless, beyond the significance of providing novel foundation models, exploring how these can be efficiently adapted to novel tasks is paramount. More particularly, *few-shot adaptation* [27], in which a small number of annotated volumes is required, is of special interest in healthcare, since each institution has limited time, budget, and particular clinical purposes, and the number of available annotated samples is usually limited. However, how to optimally adapt volumetric segmentation foundation models in this setting remains undefined, while being critical for the practical deployment of these models.

What model parameters to fine-tune? This question is essential as the choice might significantly affect the performance. Indeed, in the abundant literature on few-shot classification, most of the existing methods operate on the output embedding space [6,3,16,10], a process often referred to as linear probing. This involves fine-tuning the weights of the last linear-classifier layer while freezing the rest of the network, a common strategy in the context of few-shot segmentation [4,12]. Indeed, full fine-tuning (FFT), i.e., fine-tuning all the learnable parameters of the model, is widely avoided in the context of few-shot image classification, as it is prone to over-fitting, substantially degrading the performances [6,3]. Moreover, along with the rise of large foundation models, both in computer vision [25] and medical imaging [23,30,26,17], which involves millions or even billions of parameters, FFT has becomes less appealing in practice, as it requires substantial computational and memory resources.

Parameter-Efficient Fine-Tuning (PEFT) has emerged as an alternative to address the limitations of FFT. Recently popularized in NLP [14], and later adopted in computer vision [11,36,38], PEFT fine-tunes only a small, carefully selected [33,1] or added [15] set of the trainable parameters, making the adaptation of large models lighter (from computation and memory standpoints). Notable PEFT methods include BitFit [33], which updates only the bias terms of transformer-based models and is commonly used in NLP. Moreover, in the recent literature on vision-language models, PEFT methods have shown excellent performance in few-shot regimes. This includes the seminal work of CoOp, which pioneered prompt learning in VLMs [38], as well as Adapters [36] and, more recently, Low-Rank Adaptation (LoRA) [34]. These recent developments challenged the status quo in few-shot image classification, showing that going beyond linear probing, i.e., updating the inner representations of the models or



(a) Robustness **(b) Parameter efficiency**

Fig. 1: **Adaptive-LoRA.** We introduce a novel few-shot PEFT technique for **A**daptive **R**ank Segme**nt**a**t**ion (**ARENA**) that is (a) *robust to rank initialization* and (b) *enhances the parameter efficiency vs. performance trade-off.*

the input text prompts, could be beneficial. Given the growing success of PEFT in vision and NLP, there is an increasing interest in its application to medical domains and in the few-shot learning settings. This includes, for instance, the recent medical vision-language benchmark in [10], as well as the few-shot parameter-efficient fine-tuning framework introduced in [27] for volumetric organ segmentation, and which closely relates to the setting explored in this work.

Low-rank adaptation is a subcategory of PEFT methods, which integrates additional adaptable low-rank matrices to approximate the weight matrices during adaptation, while keeping the original model weights fixed. This approach was first introduced in the seminal work in [15] for NLP tasks, and has since inspired various extensions [29,35,9,5,8]. Following its success in NLP, LoRA has recently garnered increasing attention in computer vision, with the development of several promising approaches. A notable extension is the recent CLIP-LoRA method [34], which customized low-rank adaptation to vision-language models, showing highly competitive performances in few-shot classification. Technically, the standard LoRA baseline approximates the incremental updates ΔW of the pre-trained weights W_0 as the product of two low-rank matrices, A and B. This low-rank modification operation is defined as $W = W_0 + \Delta W = W_0 + BA$ where $A \in \mathrm{R}^{r \times n}$ and $B \in \mathrm{R}^{m \times r}$, with r representing the intrinsic rank, which is typically much smaller than m and n (i.e., $r << (m, n)$). Following this framework, only the learnable parameters of matrices A and B are optimized, while the original model parameters remain fixed. Despite its efficiency in reducing the number of trainable parameters and its good empirical performance, LoRA still operates with a fixed rank throughout the optimization process. This constraint limits its flexibility, as the optimal rank selection may vary across different downstream tasks (see Fig. 1(a)). This limitation is exacerbated in few-shot regimes, where relying on validation data for finding the optimal configuration is unrealistic.

Contributions. Building on the above-mentioned momentum in the PEFT and few-shot literature, we investigate low-rank adaptation in volumetric organ segmentation. Our contributions are: *i*) We highlight the limitations of LoRA regarding model selection in few-shot regimes, motivated by the observation in Fig. 1(a), which shows that the best rank could change significantly from one task (i.e., an organ in this case) to another; *ii*) We propose **ARENA**, an **A**daptive **R**ank Segmen**t**a**t**ion method. Viewing the low-rank representation of the trainable weight matrices as a singular value decomposition, we introduce an l_1 sparsity regularizer to the loss function, and tackle it with a proximal optimizer. The regularizer could be viewed as a penalty on the decomposition rank. Hence, its minimization enables finding task-adapted ranks automatically; and *iii*) We provide comprehensive experiments showcasing its benefits w.r.t. LoRA in a strict few-shot regime, especially when transferring foundation models to novel segmentation tasks.

2 Methods

Preliminaries. Building on the few-shot PEFT setting introduced recently in [27], this study focuses on adapting a large foundation model for volumetric organ segmentation in a realistic clinical scenario, considering the resource limitations of medical institutions. The adaptation process is designed to be both data-efficient (i.e., in a few-shot regime) and computationally lightweight (i.e., following a PEFT paradigm). The adaptation process relies on only a few labeled examples in the target conditions, a.k.a. the *support set*, to tackle real clinical settings where annotated medical data is scarce. Each few-shot segmentation task involves a support set containing fully labeled volumetric organ samples, denoted as $D_S = (X_k, Y_k)_{k=1}^{K}$, where k (typically $K \leq 16$) represents the total number of support samples used for training. It also contains a single query volume X for inference. To mitigate over-fitting in few-shot regimes and further reduce the computational overhead, we focus on PEFT strategies, in which only a tiny subset of the learnable parameters must be fine-tuned, ensuring efficient model adaptation while preserving high segmentation performance.

Proposed Adaptive Rank sEgmeNtAtion (ARENA). We consider the settings where we fine-tune a pre-trained model by optimizing a supervised-learning loss function $\mathcal{L}(W)$ using the labeled support set. In our experiments, we use the Dice loss, which is widely deployed in medical image segmentation. However, our ARENA framework is applicable to any loss function. To enable automated adjustment of the rank during the adaptation process, let us write the decomposition of the weight matrices in LoRA in the form of a singular value decomposition (SVD):

$$W = W_0 + \Delta W = W_0 + B\,\mathrm{Diag}(v)\,A, \tag{1}$$

where v is an r-dimensional vector containing the singular values, and A and B are two low-rank matrices. In SVD decomposition $B\,\mathrm{Diag}(v)\,A$, the number of non-zero elements of the vector of diagonal elements v, i.e., $\|v\|_0$, determines the

rank of the decomposition [9]. Therefore, we can control the rank by imposing a sparsity regularizer on vector v. To do so, we add the following l_1 regularizer to the loss function, as minimizing the l_1 norm promote vector sparsity:

$$\mathcal{L}(A, B, v) + \lambda \|v\|_1 \tag{2}$$

where $\mathcal{L}(A, B, v)$ is the loss function that now needs to be optimized over the blocks of learnable parameters A, B and v.

Optimization. To optimize regularized loss (2) during adaptation, we proceed with a block-coordinate descent process, which alternates two optimization steps: one fixes vector v and optimizes $\mathcal{L}$ over (A, B) via standard gradient steps, and the other fixes (A, B) and optimizes $\mathcal{L}$ over v via proximal steps (to account for the non-smooth l_1 regularizer). More precisely, the updates with respect to A and B at iteration t are defined by standard gradient steps:

$$A^{(t+1)} = A^{(t)} - \eta \nabla_A \mathcal{L}(A, B, v), \quad B^{(t+1)} = B^{(t)} - \eta \nabla_B \mathcal{L}(A, B, v), \tag{3}$$

where η denotes the learning rate, and $\nabla_A \mathcal{L}$ and $\nabla_B \mathcal{L}$ represent the gradients of the loss function with respect to A and B, respectively. Besides, each vector v is updated using a proximal update step [2], enforcing sparsity through an l_1 regularization. At each iteration t, this update is obtained by solving the following problem:

$$v^{(t+1)} = \arg\min_v \left(\frac{1}{2\eta_t} \|v - (v^{(t)} - \rho \nabla_v \mathcal{L}(A, B, v))\|_2^2 + \lambda \|v\|_1 \right). \tag{4}$$

where $\nabla_v \mathcal{L}(A, B, v)$ denotes the gradient with respect to v. The closed-form solution to this optimization problem yields the update rule for v, given by:

$$v^{(t+1)} = \text{prox}_{\eta_t \lambda \|\cdot\|_1}(v^{(t)} - \rho \nabla_v \mathcal{L}(A, B, v)) = \xi(v^{(t)} - \rho \nabla_v \mathcal{L}(A, B, v), \eta_t \lambda), \tag{5}$$

where $\xi(x, \tau)$ is the soft thresholding function defined as:

$$\xi(x, \tau) := \begin{cases} x - \tau, & x > \tau \\ 0, & -\tau \leq x \leq \tau \\ x + \tau, & x < -\tau \end{cases} \tag{6}$$

This function dynamically regulates the intrinsic rank during adaptation by eliminating small values and scaling down larger ones. It enforces structured sparsity while maintaining the model's expressive capacity.

3 Experiments

3.1 Setup

Foundation model. A 3D-SwinUNETR [13] with open-access weights from a multi-task, multi-dataset supervised pre-training in [27] is employed. Using a

supervised learning objective, this model was pre-trained on 2,048 CT scans, addressing the volumetric segmentation of 29 base anatomical structures.

Adaptation tasks. We evaluate the different methods in two types of tasks: *i*) **Base organs**: anatomical structures used during pre-training, for which the pre-trained model can perform zero-shot predictions, e.g., spleen, left kidney, gallbladder, esophagus, liver, pancreas, stomach, duodenum, and aorta. While we explore binary segmentation in TotalSegmentator dataset [31], i.e., one organ is tackled at a time, multi-class adaptation is assessed using FLARE'22 [22]. *ii*) **Novel organs**: structures unseen by the network. In particular, we focus on heart parcellation, i.e., individual segmentation of heart-myocardium (MYO), left atrium (LA), right atrium (RA), left ventricle (LV), and right ventricle (RV), from TotalSegmentator volumes. **Volume pre-processing:** CT scans are standardized following the same pipeline as in pre-training [27].

Adaptation training details. The model takes as input six patches of size $96 \times 96 \times 96$ per volume in each iteration with a batch size of one volume. AdamW is used as optimizer, and a cosine decay learning rate scheduler of 200 epochs is defined for training. The training loss is monitored in the support set, and early stopping is applied upon convergence, assuming its relative improvement over 20 epochs does not exceed 1% of the loss value at the beginning of that period.

PEFT implementation details. Following [27], the decoder is frozen when transferring to known tasks, while it is entirely updated for novel organs. Hence, PEFT methods are incorporated uniquely into the encoder. For the proposed **ARENA**, the low-rank modifications are applied to the key-value layers of each Transformer block. The matrices A and B follow the standard LoRA initialization, while the gating vector v is initialized from a uniform distribution over the interval $[-1, 1]$. During optimization, the low-rank matrices are updated using a base learning rate of 10^{-3}. The gating vector parameters are adjusted to $\lambda = 0.5$ and $\rho = 0$ according to the updating rule in Eq. 5. Across all experiments, the initial rank of ARENA is set to 8, unless otherwise specified.

Baselines. We include: *i*) black-box strategies, i.e., linear probing; *ii*) full fine-tuning of the pre-trained model (FFT), and *iii*) popular PEFT methods, i.e., selective tuning methods such as BitFit [33], affine layer normalization (Affine-LN) [1], and additive vanilla LoRA [15] and AdaLoRA [35]. We adhere to the settings provided in [27], except where explicitly stated, regarding the specific hyper-parameters. **Evaluation protocol.** The train/test splits in [27] are employed. From training, $K \in \{5, 10\}$ labeled examples are retrieved for training in the few-shot data regime, following the proposed realistic, validation-free adaptation. The test data remains fixed across different trainings, and evaluation is performed through DICE score. Results are averaged across 3 random seeds.

3.2 Results

Transferability to new tasks (Table 1). First, we assess the transfer learning capabilities of the pre-trained foundation model to segment novel structures. In this setting, FFT fails to scale properly with increasing shots ($K = 10$) by

Table 1: **Transfer performance to new tasks on TotalSegmentator.** As stated in Section 3.1, each method is combined with decoder fine-tuning.

	Method	MYO	LA	RA	LV	RV	**Avg.**
5-shot	Linear probe	51.98	38.99	40.35	53.27	31.08	43.13
	BitFit [33]	51.53	39.01	40.19	53.27	31.03	43.01
	Affine-LN [1]	51.68	38.82	40.08	**53.34**	31.06	40.41
	FFT	**52.03**	43.98	49.91	51.22	33.03	46.03
	LoRA [15]	41.83	36.53	45.67	43.42	37.05	40.90
	AdaLoRA [35]	50.28	43.59	37.35	48.53	38.73	43.70
	ARENA (*Ours*)	48.32	**51.97**	**54.38**	50.65	**43.69**	**49.80**
10-shot	Linear probe	64.50	63.47	66.86	69.12	62.60	65.31
	BitFit [33]	64.18	64.15	66.35	69.79	62.61	64.42
	Affine-LN [1]	64.39	63.62	67.95	69.93	63.66	65.91
	FFT	59.07	54.05	63.06	64.38	59.50	60.01
	LoRA [15]	60.31	65.2	78.44	64.05	65.29	66.66
	AdaLoRA [35]	61.57	68.1	60.81	59.76	55.27	61.10
	ARENA (*Ours*)	**75.29**	**81.8**	**82.93**	**74.2**	**74.82**	**77.81**

performing nearly -5.0 points below PEFT methods and linear probing. Furthermore, LoRA does not provide consistent gains w.r.t. these baselines across shots. In contrast, the proposed ARENA provides substantial improvements, with performance gains of $+8.9$ and $+11.2$ over LoRA for $K = 5$ and $K = 10$, respectively. Hence, ARENA is the best transfer learning strategy, allowing flexible feature reuse that benefits robustness when tackling challenging, novel tasks.

Transferability to base tasks in TotalSegmentator (Table 2). First, one may notice that, in contrast to the general belief, FFT is a robust alternative in the low data regime, with consistent improvements over zero-shot predictions. However, as we later discuss, this alternative requires expensive resources for transferring foundation models. PEFT methods are an appealing alternative, which fall close in performance to full fine-tuning. Notably, selective methods such as the baseline Affine-LN are strong baselines that do not introduce additional hyper-parameters nor require explicit control. In contrast, the popular LoRA depends on fine-grained model selection and fails to improve over PEFT baselines in the proposed validation-free scenario. The proposed ARENA alleviates this issue, resulting in consistent average improvements over vanilla LoRA of $+0.9$ and $+1.6$ for the two explored few-shot data regimes, surpassing linear probing, BitFit, and Affine-LN. It is worth noting that ARENA also outperforms full fine-tuning for $K = 10$, which underscores its improved data scalability.

Generalization across datasets. Table 3 demonstrates the robustness of the improvements of the proposed ARENA w.r.t. vanilla LoRA. Results in FLARE'22 for multi-class segmentation align with the observation from TotalSegmentator, with ARENA providing average gains of nearly $+1.2$ and $+0.5$.

Robustness against bad rank initialization. We now dig into the limitations of vanilla LoRA to provide satisfactory validation-free results. We focus on rank initialization, as illustrated in Fig. 1(a), and explore how the transfer performance relies on this critical choice. We perform ablation studies using several rank initializations, i.e., $r \in \{8, 32, 64\}$, with five shots for adaptation. For

Table 2: **Transfer performance to base organs on TotalSegmentator.**

	Method	Spl	lKid	Gall	Eso	Liv	Pan	Sto	Duo	Aor	**Avg.**
	Zero-shot	91.34	89.05	77.18	36.73	93.04	78.15	75.86	44.01	63.35	72.08
5-shot	Linear probe	91.42	89.51	78.15	45.98	92.69	78.31	**76.64**	62.88	69.06	76.07
	BitFit [33]	**92.00**	88.11	71.11	50.00	92.38	78.36	73.00	56.80	73.43	75.02
	Affine-LN [1]	91.43	**89.95**	74.95	**50.65**	93.04	78.81	72.41	63.05	76.71	76.78
	FFT	91.66	87.22	73.52	45.74	**93.87**	**80.34**	66.24	**67.31**	**86.18**	**76.90**
	LoRA [15]	91.48	88.65	77.93	48.02	92.81	75.80	72.95	56.15	79.43	75.91
	AdaLoRA [35]	91.28	89.15	78.11	43.76	92.98	78.35	76.17	58.42	66.49	74.96
	ARENA (*Ours*)	91.77	89.63	**79.14**	49.48	93.09	78.24	73.05	58.59	78.17	76.80
10-shot	Linear probe	91.72	**89.78**	78.49	47.01	92.16	78.14	**76.80**	63.63	69.91	76.40
	BitFit [33]	90.85	87.68	75.92	47.92	91.85	79.83	66.35	64.10	77.98	75.83
	Affine-LN [1]	89.22	87.88	73.48	51.11	91.29	80.05	65.99	62.42	82.42	75.98
	FFT	89.61	84.79	76.07	**56.82**	90.89	74.87	60.78	**71.29**	**91.81**	77.44
	LoRA [15]	89.94	89.47	80.65	46.11	92.94	**81.18**	66.41	61.76	81.66	76.68
	AdaLoRA [35]	91.28	89.27	80.01	45.72	92.90	78.44	76.32	61.81	68.19	75.99
	ARENA (*Ours*)	**92.28**	89.58	**84.49**	50.83	**93.01**	80.27	68.35	62.95	82.46	**78.25**

Table 3: **Transfer performance on an alternative dataset (FLARE'22).**

	Method	Spl	lKid	Gall	Eso	Liv	Pan	Sto	Duo	Aor	**Avg.**
5-shot	LoRA [15]	85.54	75.68	54.59	73.59	93.98	82.72	74.09	42.59	91.15	74.88
	ARENA (*Ours*)	**88.06**	**75.86**	**55.71**	**75.04**	**94.91**	**83.61**	**76.21**	**43.15**	**91.52**	**76.01**
10-shot	LoRA [15]	86.37	**77.05**	55.06	73.98	94.05	83.17	77.15	46.23	91.42	76.05
	ARENA (*Ours*)	**87.69**	76.82	**55.4**	**75.00**	**95.06**	**83.59**	**77.35**	**46.58**	**91.6**	**76.57**

example, for the aorta, the used rank ($r = 8$) is suboptimal, and transfer would benefit from more flexibility, e.g., using a rank of 32 (+1.7). In contrast, our adaptive low-rank adaptation, ARENA, is more robust to bad rank initialization, as shown in Fig. 1(a). Consequently, the overall transfer results approximate the greedy rank search in LoRA, underscoring its strength and data efficiency, particularly when adapting models in the practical low-shot regime.

Comparison to other LoRA variants. Our results demonstrate that ARENA consistently outperforms AdaLoRA [35] on both base and novel organs, yielding average gains of approximately +2 and +16 DICE points, respectively, for $k = 10$. While our method, ARENA, shares the same objective of rank adaptation with other SVD-based LoRA variants such as AdaLoRA [35] and DyLoRA [29], it differs from them in how this adaptation is achieved. These methods deploy heuristic rules and manual learning schedules [35] or rank sampling [29], whereas our method directly integrates rank adaptation into the training objective through L1 regularization, allowing the effective rank to be learned in an entirely data-driven manner, without extensive hyper-parameter tuning.

Computational efficiency. Figure 1(b) illustrates the performance/efficiency trade-off for TotalSegmentator tasks. One can readily notice that vanilla LoRA falls short in performance despite being more parameter-efficient than FFT, training 900× fewer parameters. In contrast, the proposed ARENA approaches FFT while not incurring additional computational overhead.

4 Conclusions

This work has addressed the parameter-efficient adaptation of segmentation models in few-shot regimes. Despite PEFT's computational benefits, we observe that popular strategies, such as LoRA, have particular limitations in few-shot adaptation scenarios, which are common in healthcare applications. These point out to the critical role of certain hyper-parameters in additive PEFT strategies, which control model expressiveness, and to the challenges that model selection involves in low-data regimes. The proposed adaptive low-rank strategy, ARENA, alleviates such a burden by promoting sparsity, and provides consistent gains. In this context, we anticipate that adaptive, validation-free techniques will be pivotal in future works in few-shot segmentation.

Acknowledgments. This work was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Montréal University Hospital Research Center (CRCHUM). We also thank Calcul Quebec and Compute Canada.

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

References

1. Basu, S., et al.: Strong baselines for parameter efficient few-shot fine-tuning. In: AAAI. vol. 38, pp. 11024–11031 (2024) 2, 6, 7, 8
2. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM journal on imaging sciences **2**(1), 183–202 (2009) 5
3. Boudiaf, M., Ziko, I., et al.: Information maximization for few-shot learning. Neural Information Processing Systems (NeurIPS) **33**, 2445–2457 (2020) 2
4. Boudiaf, M., et al.: Few-shot segmentation without meta-learning: A good transductive inference is all you need? In: Computer Vision and Pattern Recognition (CVPR). pp. 13979–13988 (2021) 2
5. Chavan, A., Liu, Z., Gupta, D., Xing, E., Shen, Z.: One-for-all: Generalized lora for parameter-efficient fine-tuning. arXiv preprint arXiv:2306.07967 (2023) 3
6. Chen, W.Y., Liu, Y.C., Kira, Z., Wang, Y.C.F., Huang, J.B.: A closer look at few-shot classification. In: International Conference on Learning Representations (ICLR) (2019) 2
7. Chen, X., Wang, X., Zhang, K., Fung, K.M., Thai, T.C., Moore, K., Mannel, R.S., Liu, H., Zheng, B., Qiu, Y.: Recent advances and clinical applications of deep learning in medical image analysis. Medical Image Analysis **79**, 4 (2022) 1
8. Dettmers, T., et al.: Qlora: Efficient finetuning of quantized LLMs. Neural Information Processing Systems (NeurIPS) **36**, 10088–10115 (2023) 3
9. Ding, N., et al.: Sparse low-rank adaptation of pre-trained language models. In: EMNLP (2023) 3, 5
10. Fereshteh Shakeri, Y., et al.: Few-shot adaptation of medical vision-language models. In: MICCAI (2024) 2, 3
11. Gao, P., et al.: Clip-adapter: Better vision-language models with feature adapters. International Journal of Computer Vision **132**(2), 581–595 (2024) 2

12. Hajimiri, S., Boudiaf, M., et al.: A strong baseline for generalized few-shot semantic segmentation. In: Computer Vision and Pattern Recognition (CVPR). pp. 11269–11278 (2023) 2
13. Hatamizadeh, A., et al.: Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In: Intervention Brainlesion MICCAIW. pp. 272–284 (2022) 5
14. He, J., Zhou, C., Ma, X., Berg-Kirkpatrick, T., Neubig, G.: Towards a unified view of parameter-efficient transfer learning. In: International Conference on Learning Representations (ICLR) (2022) 2
15. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al.: Lora: Low-rank adaptation of large language models. International Conference on Learning Representations (ICLR) **1**(2), 3 (2022) 2, 3, 6, 7, 8
16. Huang, Y., et al.: LP++: A surprisingly strong linear probe for few-shot CLIP. In: Computer Vision and Pattern Recognition (CVPR) (2024) 2
17. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. In: International Conference on Computer Vision (ICCV). pp. 4015–4026 (2023) 2
18. Li, W., Yuille, A., Zhou, Z.: How well do supervised models transfer to 3d image segmentation? In: International Conference on Learning Representations (ICLR). pp. 1–16 (2024) 2
19. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Medical image analysis **42**, 60–88 (2017) 1
20. Liu, J., Zhang, Y., Chen, J.N., Xiao, J., Lu, Y., Landman, B.A., Yuan, Y., Yuille, A., Tang, Y., Zhou, Z.: CLIP-Driven universal model for organ segmentation and tumor detection. In: International Conference on Computer Vision (ICCV). pp. 21152–21164 (2023) 2
21. Lu, M.Y., et al.: A visual-language foundation model for computational pathology. Nature Medicine **30**, 863–874 (2024) 2
22. Ma, J., et al.: Unleashing the strengths of unlabeled data in pan-cancer abdominal organ quantification: the flare22 challenge. Lancet Digit Health (2024) 6
23. Moor, M., Banerjee, O., et al.: Foundation models for generalist medical artificial intelligence. Nature **616**(7956), 259–265 (2023) 2
24. Moor, M., et al.: Foundation models for generalist medical artificial intelligence. Nature **616**, 259–265 (4 2023) 2
25. Radford, A., et al.: Learning transferable visual models from natural language supervision. In: ICML. pp. 8748–8763. PMLR (2021) 2
26. Silva-Rodriguez, J., et al.: A foundation language-image model of the retina (flair): Encoding expert knowledge in text supervision. Medical Image Analysis **99**, 103357 (2025) 2
27. Silva-Rodríguez, J., Dolz, J., Ben Ayed, I.: Towards foundation models and few-shot parameter-efficient fine-tuning for volumetric organ segmentation. Medical Image Analysis **103**, 103596 (2025) 2, 3, 4, 5, 6
28. Ulrich, C., et al.: Multitalent: A multi-dataset approach to medical image segmentation. In: MICCAI. pp. 508–518 (2023) 2
29. Valipour, M., et al.: Dylora: Parameter-efficient tuning of pre-trained models using dynamic search-free low-rank adaptation. In: ACL. pp. 3274–3287 (2023) 3, 8
30. Wang, Z., Wu, Z., Agarwal, D., Sun, J.: Medclip: Contrastive learning from unpaired medical images and text. In: Proceedings of EMNLP. vol. 2022, p. 3876 (2022) 2

31. Wasserthal, J., et al.: Totalsegmentator: robust segmentation of 104 anatomic structures in ct images. Radiology: Artificial Intelligence **5**(5), e230024 (2023) 2, 6
32. Wójcik, M.A.: Foundation models in healthcare: Opportunities, biases and regulatory prospects in europe. EGOVIS **13429**, 32–46 (2022) 2
33. Zaken, E., et al.: Bitfit: Simple parameter-efficient fine-tuning for transformer-based masked language models. In: ACL (2022) 2, 6, 7, 8
34. Zanella, M., Ben Ayed, I.: Low-rank few-shot adaptation of vision-language models. In: Computer Vision and Pattern Recognition (CVPR). pp. 1593–1603 (2024) 2, 3
35. Zhang, Q., et al.: Adalora: Adaptive budget allocation for parameter-efficient fine-tuning. In: International Conference on Learning Representations (ICLR) (2023) 3, 6, 7, 8
36. Zhang, R., et al.: Tip-adapter: Training-free adaption of clip for few-shot classification. In: European Conference on Computer Vision (ECCV). pp. 493–510 (2022) 2
37. Zhang, S., Metaxas, D.: On the challenges and perspectives of foundation models for medical image analysis. Medical Image Analysis **91**, 102996 (2024) 2
38. Zhou, K., Yang, J., Loy, C.C., Liu, Z.: Learning to prompt for vision-language models. International Journal of Computer Vision **130**(9), 2337–2348 (2022) 2