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

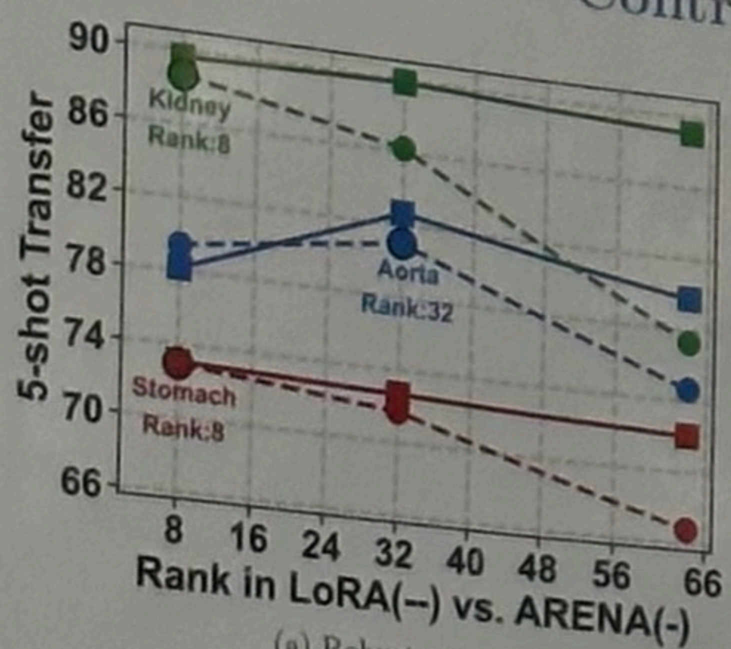
Low Rank Adaptation

- **Idea.** Freeze W_0 and approximates the incremental updates ΔW as the product of two low-rank matrices, A and B.

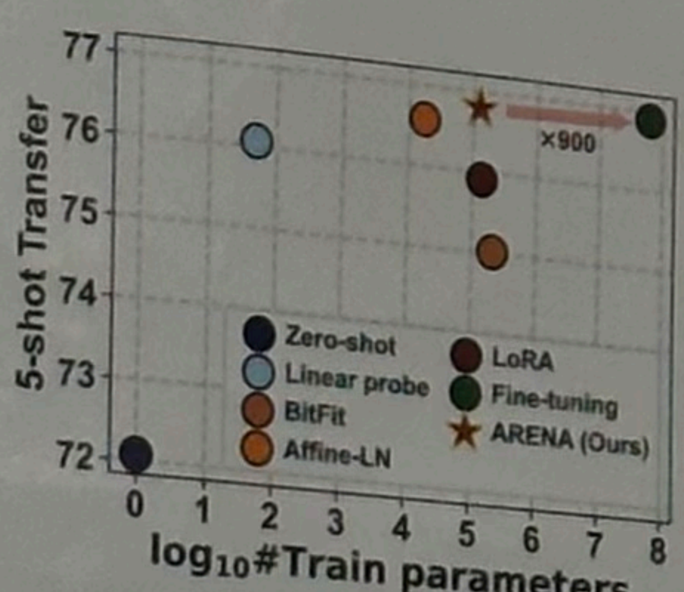
$$W = W_0 + \Delta W = W_0 + BA, \quad A \in \mathbb{R}^{r\times n}; B \in \mathbb{R}^{m\times r}; r \ll \min(m,n). \tag{1}$$

- **Limitation.** Standard LoRA operates with a fixed rank throughout optimization; The optimal rank selection often vary across different downstream tasks, limiting its flexibility.

Contributions



(a) Robustness



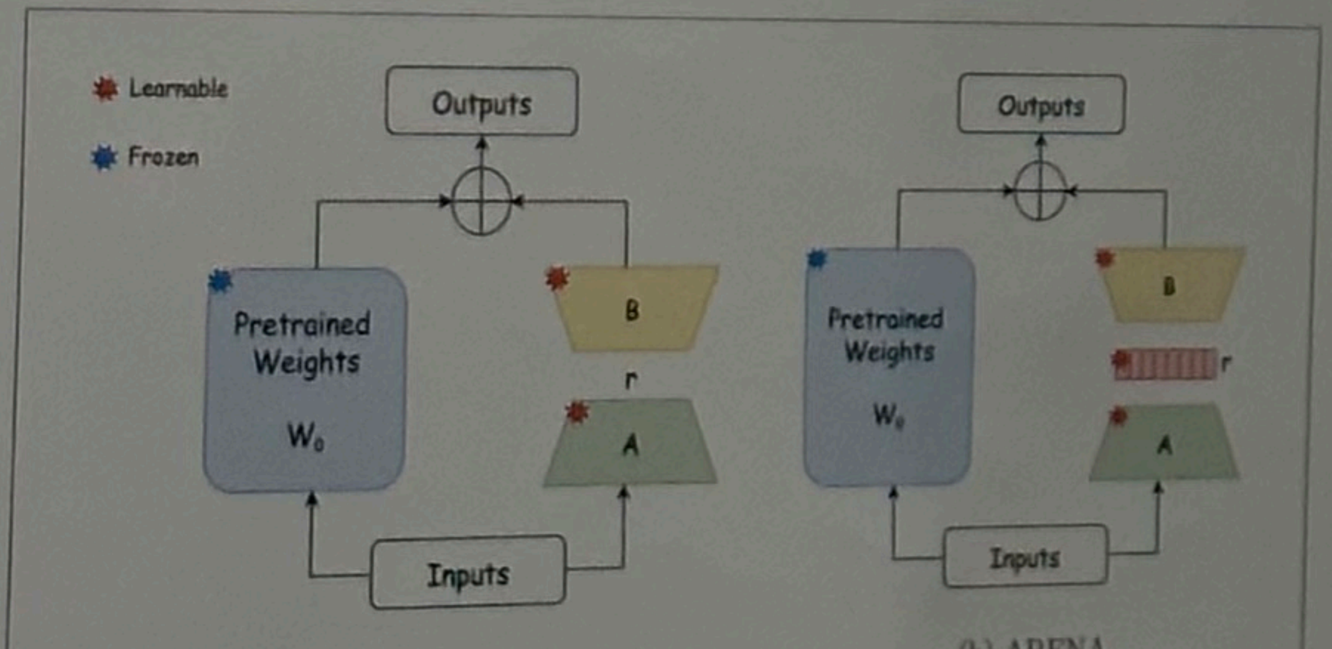
(b) Parameter efficiency

- We highlight **standard LoRA limitations** regarding **rank selection in few-shot regimes**; The optimal rank varies substantially across tasks.
- We propose **ARENA** (Adaptive Rank Segmentation), which **models low-rank updates via SVD** and integrates an l_1 **sparsity regularizer**; **Proximal updates** shrink small singular values, enabling **automatic, task-specific rank discovery**; Our framework is **loss agnostic**.
- We provide **comprehensive experiments** in **realistic clinical settings**, covering **14 organs** (**9 base** across two datasets and **5 novel**), and demonstrate consistent gains.

Methods

1. **SVD-like factorization.** We express LoRA updates in singular-value form.

$$W = W_0 + \Delta W = W_0 + B\,\mathrm{Diag}(v)\,A \tag{2}$$



(a) LoRA (b) ARENA

2. **Rank Control.** We impose an l_1 regularization on the gating vector v to promote sparsity; The effective rank is given by the number of non-zero entries in v.

$$\mathcal{L}(A,B,v) + \lambda\|v\|_1 \tag{3}$$

3. **Dynamic Adaptation – block coordinate descent.**
   - **Step 1–Update** (A, B). With v fixed, we optimize A and B via gradient descent.

$$A^{(t+1)} = A^{(t)} - \eta\nabla_A\mathcal{L}(A,B,v), \quad B^{(t+1)} = B^{(t)} - \eta\nabla_B\mathcal{L}(A,B,v) \tag{4}$$

   - **Step 2–Update** v. With (A, B) fixed, we update v via proximal steps for the l_1 penalty.

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

where $\xi(.,\tau)$ is the soft thresholding operator

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

Experimental Results

- **Transferability to new tasks in TotalSegmentator.** 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 [3]	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 [2]	41.83	36.53	45.67	43.42	37.05	40.90
	AdaLoRA [4]	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 [3]	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 [2]	60.31	65.2	78.44	64.05	65.29	66.66
	AdaLoRA [4]	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**

- **Transferability to base tasks in TotalSegmentator.** Average is reported over 9 organs (Spl, lKid, Gall, Eso, Liv, Pan, Sto, Duo, Aor).

	Method	Gall	Eso	Liv	Pan	Aor	Avg.
	Zero-shot	77.18	36.73	93.04	78.15	63.35	72.08
5-shot	Linear probe	78.15	45.98	92.69	78.31	69.06	76.07
	BitFit [3]	71.11	50.00	92.38	78.36	73.43	75.02
	Affine-LN [1]	74.95	**50.65**	93.04	78.81	76.71	76.78
	FFT	73.52	45.74	**93.87**	**80.34**	**86.18**	**76.90**
	LoRA [2]	77.93	48.02	92.81	75.80	79.43	75.91
	AdaLoRA [4]	78.11	43.76	92.98	78.35	66.49	74.96
	ARENA (*Ours*)	**79.14**	49.48	93.09	78.24	78.17	76.80
10-shot	Linear probe	78.49	47.01	92.16	78.14	69.91	76.40
	BitFit [3]	75.92	47.92	91.85	79.83	77.98	75.83
	Affine-LN [1]	73.48	51.11	91.29	80.05	82.42	75.98
	FFT	76.07	**56.82**	90.89	74.87	**91.81**	77.44
	LoRA [2]	80.65	46.11	92.94	**81.18**	81.66	76.68
	AdaLoRA [4]	80.01	45.72	92.90	78.44	68.19	75.99
	ARENA (*Ours*)	**84.49**	50.83	**93.01**	80.27	82.46	**78.25**

- **Generalization across datasets(FLARE'22).** Average is reported over 9 organs (Spl, lKid, Gall, Eso, Liv, Pan, Sto, Duo, Aor).

	Method	Gall	Eso	Liv	Pan	Aor	Avg.
5-shot	LoRA [2]	54.59	73.59	93.98	82.72	91.15	74.88
	ARENA (*Ours*)	**55.71**	**75.04**	**94.91**	**83.61**	**91.52**	**76.01**
10-shot	LoRA [2]	55.06	73.98	94.05	83.17	91.42	76.05
	ARENA (*Ours*)	**55.4**	**75.00**	**95.06**	**83.59**	**91.6**	**76.57**

References

[1] Basu, S., et al.: Strong baselines for parameter efficient few-shot fine-tuning. In: AAAI vol. 38, pp. 11024–11031 (2024)
[2] 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)
[3] Zaken, E., et al.: Bitfit: Simple parameter-efficient fine-tuning for transformer-based masked language models. In: ACL (2022)
[4] Zhang, Q., et al.: Adalora: Adaptive budget allocation for parameter-efficient fine-tuning. In: International Conference on Learning Representations (ICLR) (2023)

Feel Free To use Our ARENA!