



This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

PriLoRA: Prior-Conditioned Low-Rank Adapters for Medical Vision Models

Amirhossein Kazerooni^{1,2,3†*}, Sanaz Karimijafarbigloo^{4†}, Ali Aghayari^{5†}, Ali Ghasemzadeh Khorzoughi^{5†}, Reza Azad⁴, Wojciech Samek^{6,7}, Dorit Merhof⁴, Michael Brudno^{1,2,3}, and Babak Taati^{1,2,3}

¹University of Toronto, ²Vector Institute, ³University Health Network, ⁴University of Regensburg, ⁵Independent Researcher, ⁶Technische Universität Berlin ⁷Fraunhofer Heinrich Hertz Institute, amirhossein@cs.toronto.edu

Abstract. Adapting large foundation models to medical imaging is costly due to limited curated data, dense expert annotation requirements, and the expense of full fine-tuning. Parameter-efficient fine-tuning (PEFT) methods such as LoRA mitigate these costs, but linear low-rank updates can be expressivity-limited. Recent sinusoidal variants such as SineLoRA improve capacity, yet depend on a fixed frequency hyperparameter that is difficult to transfer across modalities and tasks. We propose **Prior-Conditioned Low-Rank Adapters (PriLoRA)**, an input-conditioned sinusoidal adapter that leverages input-derived prior features to modulate the adapter response per block and per sample while preserving LoRA’s modularity and parameter efficiency. PriLoRA predicts lightweight modulation parameters that control the sinusoidal update, enabling adaptive expressivity without increasing the adapter rank or requiring task-specific frequency search. We evaluate PriLoRA across multiple medical backbones, modalities, and tasks, including generation, classification, segmentation, and CT reconstruction, and observe consistent improvements in robustness and transfer over standard LoRA and representative PEFT baselines across datasets and modalities. **Code:** <https://github.com/xmindflow/PriLoRA>

Keywords: PEFT · LoRA · Medical imaging · Prior-conditioned adapters.

1 Introduction

Foundation models [23,19,18] are increasingly adopted in medical imaging due to their strong transferability. However, full fine-tuning is often expensive in memory, compute, and data, a challenge amplified by limited access to curated datasets and expert annotations [12]. This has motivated parameter-efficient fine-tuning (PEFT) methods that reuse pretrained backbones while updating only a small set of parameters [8]. LoRA [10] instantiates this idea by introducing lightweight trainable updates to selected layers while freezing the backbone,

* Corresponding author.

† Equal contribution.

substantially reducing training cost—an advantage that is particularly important in medical imaging with limited annotations and constrained resources [6,29].

Recent work in PEFT has moved beyond *vanilla* LoRA by modifying *where* and *how* lightweight updates are parameterized. Following [16], existing extensions can be broadly grouped into: (i) **adaptation-improving** designs that redistribute adapter capacity across layers or refine the update parameterization [30,14,15,17]; (ii) **operator-aware** designs that generalize LoRA to non-linear-layer structures such as convolutional kernels, enabling parameter-efficient tuning of CNN backbones for medical tasks [21]; and (iii) **efficiency-focused** designs that combine LoRA with quantization to enable low-resource fine-tuning and deployment, including healthcare-oriented decision support pipelines [4,1]. Beyond these design choices, LoRA also faces a *rank bottleneck*: purely linear low-rank updates are constrained in representational capacity by construction. SineLoRA [11] addresses this bottleneck by injecting a sinusoidal non-linearity into a low-rank parameterization, increasing the effective rank of the induced transformation. However, its performance is sensitive to the frequency choice, which is treated as a hyperparameter and can impact optimization stability [11].

In this work, we aim to retain the rank-enhancing benefit of sinusoidal low-rank parameterizations while avoiding reliance on a globally fixed frequency choice. This is particularly important in medical imaging, where different modalities and tasks demand different levels of expressivity, making a single fixed setting difficult to transfer across datasets and use cases. More broadly, this reflects a central challenge in parameter-efficient adaptation: achieving high accuracy with sufficient expressivity while preserving the desirable low-rank structure of adapter updates. We propose **Prior-Conditioned Low-Rank Adapters (PriLoRA)**, a parameter-efficient adapter that conditions a sinusoidal low-rank *weight update* on input-derived prior knowledge. PriLoRA introduces a lightweight **modulator** that takes prior features extracted from the input and predicts a small set of parameters that specialize the adapter for each block and each sample. For a linear layer with frozen weight W_0 and LoRA factors B, A , we first form the low-rank update BA , and then apply an input-conditioned sinusoidal transform: $\Delta W_{\text{Pri}} = \alpha \sin(\beta(\omega_0 BA) + \gamma) + \delta$, where α controls the amplitude, β scales the effective frequency, γ sets the phase, and δ provides an additive shift (all broadcast to the shape of BA). This turns sinusoidal LoRA into an input-conditioned adapter that adjusts its non-linear response per block and per sample, while preserving LoRA-style modularity and parameter efficiency.

2 Method

Background LoRA [10] enables parameter-efficient fine-tuning by freezing pre-trained weights and learning a low-rank residual update. For a linear layer with weight $W_0 \in \mathbb{R}^{d \times k}$, LoRA parameterizes the update as

$$\Delta W = BA, \quad B \in \mathbb{R}^{d \times r}, A \in \mathbb{R}^{r \times k}, r \ll \min(d, k), \quad (1)$$

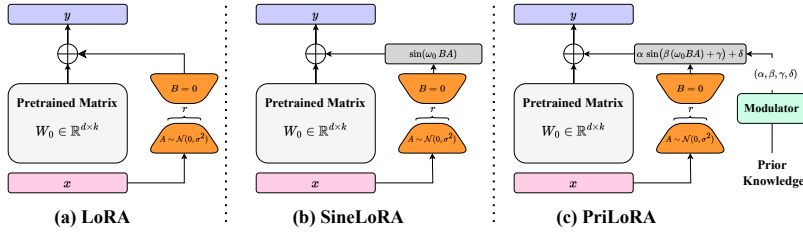


Fig. 1. Comparison of LoRA, SineLoRA, and PriLoRA. (a) *LoRA* adapts a frozen pretrained weight W_0 using a low-rank update $\Delta W = BA$ (with rank r), producing the output y via the residual path. (b) *SineLoRA* increases expressivity by applying an element-wise sinusoidal transform to the low-rank update, e.g., $\sin(\omega_0 BA)$. (c) *PriLoRA* conditions the sinusoidal update on input-derived *prior knowledge* using a lightweight modulator that predicts $(\alpha, \beta, \gamma, \delta)$, where β rescales a fixed base frequency ω_0 to obtain an input-conditioned effective frequency.

and applies the adapted layer as $y = (W_0 + \Delta W)x = W_0x + BAx$. Since only A and B are trained, LoRA reduces the number of trainable parameters from $\mathcal{O}(dk)$ to $\mathcal{O}(r(d + k))$, enabling efficient adaptation of large backbones. A central limitation of this parameterization is a *rank bottleneck*: by construction, $\text{rank}(\Delta W) \leq r$. Increasing r improves expressivity and can better approximate the generally full-rank update required for optimal adaptation, but it also increases parameter count and compute. This trade-off motivates techniques that enhance the representational capacity of low-rank updates without sacrificing parameter efficiency. SineLoRA [11] alleviates this limitation by injecting a sinusoidal non-linearity into the low-rank update. Concretely, it applies a non-linearity to a low-rank matrix, and specializes to LoRA via: $\Delta W = \frac{1}{g} \sin(\omega_0 \cdot BA)$, where $g > 0$ is a fixed gain (scaling) factor used to control the magnitude of the update and stabilize optimization. This formulation increases the effective rank and improves approximation quality without adding trainable parameters [11]. However, performance is sensitive to the hyperparameters, specifically frequency ω_0 , which is a hyperparameter that can affect optimization stability and transfer across medical tasks and modalities. This motivates an adaptive alternative that conditions the sinusoidal update on input-derived priors.

2.1 Prior-Conditioned Low-Rank Adapters (PriLoRA)

We propose **PriLoRA**, a parameter-efficient adapter that preserves the expressivity gains of sinusoidal low-rank updates while avoiding a single globally fixed effective sinusoidal behavior. SineLoRA can be viewed as applying a *fixed* sinusoidal transform to a low-rank update, where the primary controllable knob is the frequency ω_0 . Varying ω_0 exposes a rank-capacity trade-off: increasing frequency reduces redundancy and diversifies directions, leading to higher *effective rank*, but overly large values can push the update toward an unnecessarily high-rank

regime, resulting in over-parameterization and degraded generalization. In medical imaging, the appropriate operating point is not universal; different modalities and tasks induce different distribution shifts and require different adaptation capacity, making a single fixed ω_0 brittle to tune and hard to transfer.

SineLoRA fixes remaining degrees of freedom through a shared gain $\frac{1}{g}$ and zero phase/offset, enforcing one global sinusoid for all inputs. However, stable transfer depends not only on frequency but also amplitude and where the nonlinearity acts (phase and shift), which vary across medical tasks. In segmentation, a fixed sinusoid may behave nearly linear in some layers or overly saturate others, harming boundary refinement. In image generation/restoration it can underfit fine textures or introduce oscillatory artifacts, and in classification excessive nonlinearity may amplify spurious high-frequency features and reduce robustness. These task-dependent effects motivate conditioning amplitude, frequency, phase, and shift on input-derived priors instead of using a single global sinusoid.

Formulation. Using the LoRA low-rank update BA from Eq. (1), PriLoRA replaces the globally fixed sinusoid with a prior-conditioned sinusoidal transform whose amplitude, frequency scaling, phase, and offset are predicted from input-derived prior knowledge. Let $z = p(x) \in \mathbb{R}^r$ be a vector of prior features extracted from the current medical input x . A lightweight modulator $g(\cdot; \theta) : \mathbb{R}^r \rightarrow \mathbb{R}^4$ maps z to four scalars: $(\alpha, \beta, \gamma, \delta) = g(z; \theta)$, which parameterize the sinusoidal update for that input. PriLoRA then defines the weight update as

$$\Delta W_{\text{Pri}} = \alpha \sin\left(\beta (\omega_0 BA) + \gamma\right) + \delta, \quad (2)$$

where $\sin(\cdot)$ is applied element-wise and the scalar parameters are broadcast over the entries of BA . The adapted layer becomes

$$y = (W_0 + \Delta W_{\text{Pri}})x = W_0x + \Delta W_{\text{Pri}}x. \quad (3)$$

Effect of parameters. The parameters in Eq. (2) control complementary aspects of the update. The amplitude term α gates the overall strength of the adapter, scaling the deviation from W_0 and allowing PriLoRA to attenuate or amplify the update based on the difficulty or reliability of the input-derived prior (e.g., down-weighting noisy priors). The frequency-scaling term β rescales the argument of the sinusoid, effectively stretching or compressing $\sin(\cdot)$ along the BA directions: larger β induces more oscillatory behavior and higher effective rank (capturing fine-grained variations), while smaller β keeps the update closer to a low-rank, near-linear regime, which can improve stability and reduce overfitting. The phase term γ shifts where BA lies on the sinusoid, allowing PriLoRA to reposition the operating point into regions with favorable slope (e.g., avoiding saturation for certain modalities or intensity ranges) without changing the magnitude of BA itself. Finally, the offset term δ adds a learned bias to the sinusoidal output, enabling modest input-dependent baseline corrections (e.g., domain-specific shifts) without requiring a separate bias adapter. Together, these four coefficients let PriLoRA adapt the shape and strength of the sinusoidal low-rank update to the medical input, and SineLoRA is recovered as a special case by fixing $(\alpha, \beta, \gamma, \delta)$ to global constants.

Table 1. Comparison across Generation, Segmentation, Classification, and CT Reconstruction tasks. For generation, lower is better for FID and higher is better for CLIP Score, reported as mean $\pm$ std. For segmentation and classification, higher is better for all metrics. For CT reconstruction, higher PSNR/SSIM and lower LPIPS indicate better quality; results are reported at projection angles $\{50, 100, 150, 200\}$. Best performance is shown in **bold**, and second-best performance is underlined.

(a) Comparison on generation and segmentation tasks.									
Methods	Generation				Segmentation				
	FID _{Total} ↓	FID _{Test} ↓	CLIPScore ↑	# Params (M)	IoU ↑	Dice ↑	# Params (M)		
AdaLoRA [30]	151.26 ± 1.82	158.11 ± 1.65	28.08 ± 0.04	4.948	0.8209	0.8930	16.911		
DoRA [14]	114.72 ± 1.03	120.57 ± 1.15	28.00 ± 0.05	5.021	0.8110	0.8864	16.975		
LoRA [9]	112.32 ± 0.63	118.52 ± 0.63	<u>28.01 ± 0.07</u>	4.706	0.8388	0.9051	16.911		
PiSSA [17]	119.01 ± 1.56	125.19 ± 1.52	27.26 ± 0.08	4.706	<u>0.8404</u>	<u>0.9057</u>	16.911		
SineDoRA [11]	<u>79.07 ± 0.49</u>	<u>90.74 ± 0.44</u>	27.55 ± 0.13	5.021	0.7841	0.8680	16.975		
SineLoRA [11]	85.05 ± 1.43	93.84 ± 1.50	27.47 ± 0.04	4.706	0.8349	0.9029	16.911		
PriLoRA (Ours)	70.15 ± 1.23	80.73 ± 0.99	27.65 ± 0.07	4.949	0.8533	0.9156	17.856		

(b) Comparison on Classification and CT Reconstruction tasks.											
Methods	Classification					CT Reconstruction					
	AUC	ROC	Recall ↑	Precision ↑	F1-score ↑	# Trainable (M)	PSNR ↑ / LPIPS ↓ / SSIM ↑				# Trainable
							50 views	100 views	150 views	200 views	
Baseline	0.9872	0.8900	0.9131	0.8883	17.927	25.53 / 0.249 / 0.548	27.41 / 0.192 / 0.692	27.54 / 0.200 / 0.685	28.55 / 0.178 / 0.747	198.401	
AdaLoRA [30]	<u>0.9891</u>	0.8660	0.9033	0.8624	3.815	25.22 / 0.266 / 0.543	26.50 / 0.229 / 0.648	26.47 / 0.227 / 0.638	27.38 / 0.197 / 0.700	49.248	
DoRA [14]	0.9929	<u>0.8850</u>	<u>0.9151</u>	<u>0.8851</u>	3.819	25.36 / 0.257 / 0.547	<u>26.82 / 0.221 / 0.663</u>	27.17 / 0.209 / 0.669	<u>27.64 / 0.193 / 0.716</u>	49.920	
LoRA [9]	0.9803	0.8160	0.8872	0.8019	3.814	25.36 / 0.258 / 0.546	26.83 / 0.220 / 0.662	27.09 / 0.209 / 0.667	<u>27.66 / 0.193 / 0.715</u>	49.152	
PiSSA [17]	0.9768	0.8050	0.8826	0.8072	3.814	25.32 / 0.259 / 0.545	26.74 / 0.221 / 0.659	26.97 / 0.212 / 0.661	27.55 / 0.193 / 0.711	49.152	
SineDoRA [11]	0.9941	0.8780	0.9118	0.8810	3.819	<u>25.40 / 0.249 / 0.554</u>	26.69 / 0.222 / 0.663	27.12 / 0.207 / 0.677	27.24 / 0.202 / 0.696	49.920	
SineLoRA [11]	0.9849	0.8210	0.8780	0.8157	3.814	25.42 / 0.255 / 0.552	26.82 / 0.224 / 0.661	27.18 / 0.206 / 0.671	27.34 / 0.198 / 0.703	49.152	
PriLoRA (Ours)	0.9823	0.8980	0.9161	0.8989	3.876	25.48 / 0.247 / 0.567	27.53 / 0.190 / 0.693	27.83 / 0.194 / 0.693	27.85 / 0.190 / 0.720	49.740	

3 Experimental Results

We comprehensively evaluate PriLoRA across four medical vision tasks: image generation, segmentation, classification, and CT reconstruction. We report both quantitative and qualitative results across diverse medical datasets, backbone architectures, and imaging modalities, showing that conditioning on input-derived priors consistently outperforms existing low-rank adapters. Following prior work, we use a rank of 16 by default. AdaLoRA halves the rank mid-training, and we use ranks of 32 for CT reconstruction.

3.1 Generation

Dataset. We use the IU X-Ray dataset [27], which contains 3,955 radiology reports paired with frontal and lateral chest X-ray images. The data are split into 80%/16%/4% for training, validation, and testing, respectively.

Architecture. We build on Stable Diffusion v1.5 [22] using the diffusers framework, freezing the VAE and pretrained backbones while applying parameter-efficient adaptation to the UNet and CLIP text encoder [20]. PriLoRA partitions the UNet into ten blocks, each equipped with a lightweight modulation module. For each block, a prior feature vector is formed by concatenating channel-wise statistics (mean, standard deviation, min, max) of the preceding block’s hidden states with mean-pooled text embeddings to keep it efficient. A two-layer MLP with LayerNorm and SiLU predicts four block-specific parameters that

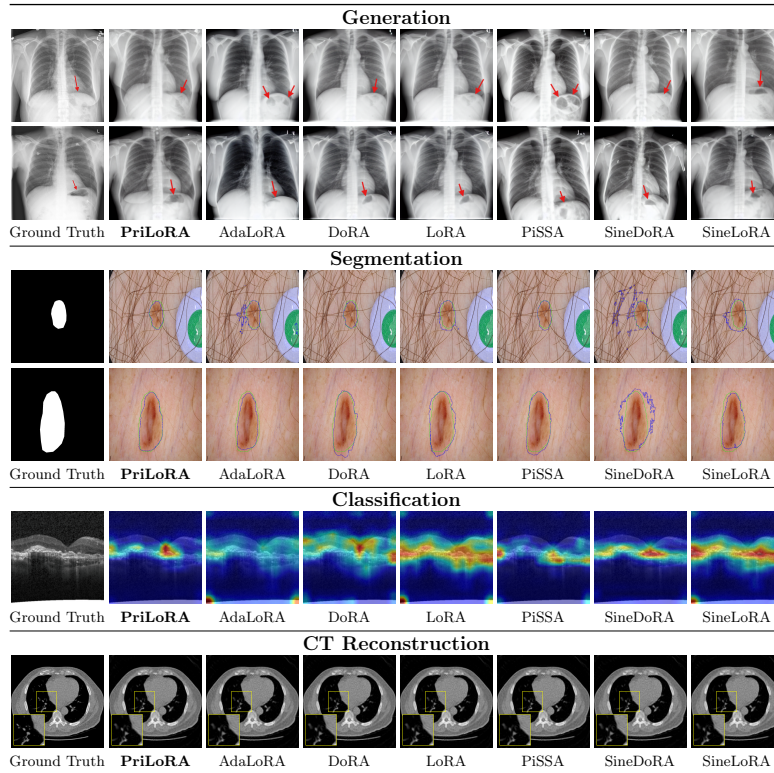


Fig. 2. Qualitative comparison across tasks. The first two rows show generation results for two prompts: (1) *There are no focal areas of consolidation. No pleural effusions. No pneumothorax. Heart size within normal limits. Osseous structures intact.* and (2) *The heart is normal in size and contour. The lungs are clear, without evidence of infiltrate. There is no pneumothorax or effusion.* The third and fourth rows present segmentation results. The fifth row shows Grad-CAM++ [2] results for Choroidal Neovascularization (ChN). The last row reports CT reconstruction results.

condition the subsequent block. Training follows the standard diffusion noise-prediction objective with MSE loss, using AdamW (learning rate (1×10^{-4})), batch size 8, and 8 epochs. Images are resized and center-cropped to (224×224) . Sampling uses 50 diffusion steps with classifier-free guidance (scale 7.5).

Analysis. We evaluate generation using FID_{Total} , FID_{Test} , and CLIPScore, where FID_{Total} compares generated samples to the full dataset and FID_{Test} to the held-out test set. Fig. 2 presents qualitative results for two representative prompts. While all methods produce visually plausible radiographs, PriLoRA more consistently preserves anatomical structure and spatial coherence, yielding images closer to the ground truth. This trend is reflected in Tab. 1, where PriLoRA achieves the lowest FID while maintaining competitive CLIP scores, indicating improved image fidelity without sacrificing text-image alignment.

3.2 Segmentation

Dataset. We use ISIC 2016 [7] for 2D melanoma segmentation, comprising 900 training images with expert lesion masks and 379 test dermoscopic images.

Architecture. We follow [26] for 2D medical segmentation and adapt Segment Anything (SAM) [13] to the medical domain while preserving its architecture. As in the generation setting, modulation uses prior features extracted via adaptive average pooling that halves the dimension of each preceding layer. The pooled features are passed through a lightweight MLP with a single 32-unit hidden layer, LayerNorm, and SiLU activation to predict the modulation parameters.

Analysis. As shown in Tab. 1, PriLoRA achieves the highest IoU and Dice, indicating improved region overlap and spatial structure preservation. Qualitative examples in Fig. 2 further show that PriLoRA produces smoother, more contiguous lesion boundaries that closely match the ground truth, while other adapters often exhibit boundary fragmentation or spurious activations. These results indicate that conditioning low-rank updates on input-derived priors leads to more stable and precise segmentation without increasing model capacity.

3.3 Classification

Dataset. We use OCTMNISTv2 [28], a four-class retinal OCT dataset (ChN, DME, Drusen, Normal) commonly used for retinal disease diagnosis [5]. We follow the official split: the training set contains 33,484 ChN, 10,213 DME, 7,754 Drusen, and 46,026 Normal images, while the validation and test sets include 2,708 and 250 samples per class, respectively.

Architecture. We extend layer-wise adaptation to the convolutional domain by attaching a lightweight modulator to each of VGG-16 [25]’s 13 convolutional layers while jointly training the two classification heads. The modulator pools global spatial and channel context and uses a single-layer MLP to predict depth-specific modulation parameters.

Analysis. We analyze attention using Grad-CAM++ [2] on the final VGG-16 convolutional layer. Compared to baseline and other PEFT methods, PriLoRA yields more localized, anatomically relevant activations, consistent with its improved precision, recall, and F1-score in Tab. 1.

3.4 CT Reconstruction

Dataset. We evaluate CT reconstruction using 256×256 lung CT images from the publicly available Kaggle Lung Nodule Analysis dataset [3].

Architecture. We adopt SIREN [24] as the base implicit neural representation for CT reconstruction, implemented as a 3-hidden-layer MLP mapping 2D coordinates (x, y) to CT intensities. Given a specified number of projection views, sinograms are generated via the Radon transform, and training minimizes the discrepancy between predicted and ground-truth sinograms by applying the Radon transform to the reconstructed output. As a baseline, SIREN is fully trained for 1,000 iterations, with an intermediate checkpoint saved at iteration

500; starting from this checkpoint, PEFT is performed for the remaining 500 iterations under the same objective. We apply PriLoRA, which modulates the hidden linear layers of SIREN while keeping the base weights frozen, using a lightweight, zero-initialized modulator (LayerNorm–SiLU–Linear) that takes a 32-dimensional pooled feature representation and predicts four modulation parameters per hidden layer to control the sinusoidal low-rank update.

Analysis. For CT reconstruction, PriLoRA achieves improved quantitative (Tab. 1) performance across PSNR, LPIPS, and SSIM at all projection counts. Qualitatively (Fig. 2), while differences are subtle, PriLoRA preserves slightly finer structural details without introducing additional artifacts.

3.5 Ablation Studies

Parameter-wise Contribution Analysis. We ablate the modulation parameters by selectively disabling each component. As shown in Tab. 2 (left), removing any parameter degrades generation quality, indicating that the components play complementary roles. Activating all four yields the best FID on both training and test sets, while partial configurations may improve individual metrics (e.g., CLIP) at the cost of overall generative fidelity.

Rank Ablation on Classification. We compare LoRA, SineLoRA, and PriLoRA across adapter ranks to assess classification performance. As shown in Tab. 2 (right), PriLoRA consistently matches or outperforms baselines, with particularly strong performance at low and mid ranks. Notably, PriLoRA remains competitive even at rank 2 and achieves peak accuracy and F1 at modest ranks, demonstrating efficient and stable adaptation without large adapter capacity.

Model Efficiency. Using CUDA event timing over 50 samples, LoRA achieves 2.40 ms inference latency, while SineLoRA, SineDoRA, and PriLoRA incur slightly higher latencies of 2.84 ms, 3.05 ms, and 2.88 ms, respectively, indicating only a modest overhead relative to LoRA while preserving parameter-efficient inference. Compared to LoRA, PriLoRA increases the number of trainable parameters by 1–6%. Compared to SineDoRA, PriLoRA uses fewer parameters for generation (−1.4%) and CT reconstruction (−0.36%), and incurs only a modest increase for segmentation (+5.2%, 0.88M) and classification (+1.5%, 0.06M).

4 Discussion and Conclusion

We introduce PriLoRA, a prior-conditioned sinusoidal low-rank adapter that improves the expressivity of parameter-efficient fine-tuning while preserving LoRA’s efficiency. By conditioning sinusoidal low-rank updates on input-derived priors, PriLoRA adapts its non-linear behavior across medical tasks without increasing adapter rank or requiring task-specific tuning. Experiments across generation, classification, segmentation, and CT reconstruction show consistent gains in robustness and transfer over standard LoRA and representative PEFT baselines.

Limitations. PriLoRA is particularly effective in settings with heterogeneous adaptation demands, limited adapter rank, and distribution shifts across modalities, where fixed hyperparameter adapters can be brittle. While the method

Table 2. Ablation studies on PriLoRA. (*Left*) parameter ablation over modulation parameters α - δ on the generation task. (*Right*) Rank ablation comparing LoRA, SineLoRA, and PriLoRA on classification, reporting AUC-ROC/accuracy (top) and precision/F1-score (bottom) for each rank.

α	β	γ	δ	FID _T	FID _{Te}	CLIP
✓	✗	✗	✗	85.51	94.88	27.45
✗	✓	✗	✗	82.20	92.21	27.56
✗	✗	✓	✗	89.18	99.05	27.96
✗	✗	✗	✓	85.57	95.32	27.50
✓	✓	✗	✗	97.48	106.79	28.16
✓	✓	✓	✗	94.47	104.96	26.78
✓	✗	✓	✓	98.02	109.25	27.50
✗	✓	✓	✓	90.03	100.86	27.90
✓	✓	✓	✓	70.15	80.73	27.65

Rank	LoRA	SineLoRA	PriLoRA
2	<u>0.9833</u> / <u>0.851</u> 0.8989 / <u>0.8480</u>	0.9858 / 0.8470 0.8985 / 0.8454	0.9754 / 0.8600 <u>0.8986</u> / 0.8586
4	0.9863 / <u>0.868</u> 0.9078 / <u>0.8644</u>	<u>0.9891</u> / 0.835 0.8935 / 0.8344	0.9906 / 0.8770 <u>0.9059</u> / 0.8745
8	<u>0.9847</u> / <u>0.842</u> <u>0.8912</u> / <u>0.8363</u>	0.9782 / 0.810 0.8820 / 0.8061	0.9864 / 0.880 0.9101 / 0.8799
16	0.9803 / 0.8160 <u>0.8872</u> / 0.8019	0.9849 / <u>0.8210</u> 0.8780 / <u>0.8157</u>	<u>0.9823</u> / 0.8980 0.9161 / 0.8989
32	<u>0.9808</u> / <u>0.870</u> <u>0.9079</u> / <u>0.8682</u>	0.9889 / 0.865 0.9058 / 0.8629	0.9803 / 0.878 0.9109 / 0.8751

adaptively selects its modulation parameters, performance may be sensitive to to extreme initialization of the base weights W_0 , which can reduce the effective adaptation capacity during training. Although we evaluate across tasks and modalities, each task is evaluated on one main dataset; multi-institution or scanner robustness remains future work. Overall, these results suggest that input-conditioned modulation is a promising direction for scalable and reliable adaptation of medical foundation models.

Acknowledgment: This work was supported by the German Research Foundation (DFG) as research unit DeSBI [KI-FOR 5363] (459422098).

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

References

1. Ansari, M.S., Khan, M.S.A., Revankar, S., Varma, A., Mokhade, A.S.: Lightweight clinical decision support system using qlora-fine-tuned llms and retrieval-augmented generation. arXiv preprint arXiv:2505.03406 (2025)
2. Bany Muhammad, M., Yeasin, M.: Eigen-cam: Visual explanations for deep convolutional neural networks. SN Computer Science **2**(1), 47 (2021)
3. Brag, Y.E.: Lung nodules detection dataset annotations. <https://www.kaggle.com/datasets/younesselbrag/lung-nodules-detection-dataset-annotations> (2025), accessed: 2026-02-12
4. Dettmers, T., Pagnoni, A., Holtzman, A., Zettlemoyer, L.: Qlora: Efficient fine-tuning of quantized llms. Advances in neural information processing systems **36**, 10088–10115 (2023)
5. Ferrara, N.: Vascular endothelial growth factor and age-related macular degeneration: from basic science to therapy. Nature medicine **16**(10), 1107–1111 (2010)
6. Guan, H., Liu, M.: Domain adaptation for medical image analysis: a survey. IEEE Transactions on Biomedical Engineering **69**(3), 1173–1185 (2021)
7. Gutman, D., Codella, N.C., Celebi, E., Helba, B., Marchetti, M., Mishra, N., Halpern, A.: Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (isbi) 2016, hosted by the international skin imaging collaboration (isic). arXiv preprint arXiv:1605.01397 (2016)

8. Han, Z., Gao, C., Liu, J., Zhang, J., Zhang, S.Q.: Parameter-efficient fine-tuning for large models: A comprehensive survey. arXiv preprint arXiv:2403.14608 (2024)
9. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: Lora: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685 (2021)
10. 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. ICLR **1**(2), 3 (2022)
11. Ji, Y., Saratchandran, H., Gordon, C., Zhang, Z., Lucey, S.: Efficient learning with sine-activated low-rank matrices. In: The Thirteenth International Conference on Learning Representations (2025), <https://openreview.net/forum?id=cWGCkd7mCp>
12. Kazerouni, A., Aghdam, E.K., Heidari, M., Azad, R., Fayyaz, M., Hacıhaliloglu, I., Merhof, D.: Diffusion models in medical imaging: A comprehensive survey. Medical image analysis **88**, 102846 (2023)
13. 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: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4015–4026 (2023)
14. Liu, S.Y., Wang, C.Y., Yin, H., Molchanov, P., Wang, Y.C.F., Cheng, K.T., Chen, M.H.: Dora: Weight-decomposed low-rank adaptation. In: Forty-first International Conference on Machine Learning (2024)
15. Mao, Y., Li, H., Pang, W., Papanastasiou, G., Yang, G., Wang, C.: Selora: Self-expanding low-rank adaptation of latent diffusion model for medical image synthesis. arXiv preprint arXiv:2408.07196 (2024)
16. Mao, Y., Ge, Y., Fan, Y., Xu, W., Mi, Y., Hu, Z., Gao, Y.: A survey on lora of large language models. Frontiers of Computer Science **19**(7), 197605 (2025)
17. Meng, F., Wang, Z., Zhang, M.: Pissa: Principal singular values and singular vectors adaptation of large language models. Advances in Neural Information Processing Systems **37**, 121038–121072 (2024)
18. Moor, M., Banerjee, O., Abad, Z.S.H., Krumholz, H.M., Leskovec, J., Topol, E.J., Rajpurkar, P.: Foundation models for generalist medical artificial intelligence. Nature **616**(7956), 259–265 (2023)
19. Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., Penna, J., Rombach, R.: Sdxl: Improving latent diffusion models for high-resolution image synthesis. arXiv preprint arXiv:2307.01952 (2023)
20. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: International conference on machine learning. pp. 8748–8763. PmLR (2021)
21. Rivera, M., Hoyos, A.: Colora: Efficient fine-tuning for convolutional models with a study case on optical coherence tomography image classification. arXiv preprint arXiv:2505.18315 (2025)
22. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10684–10695 (2022)
23. Siméoni, O., Vo, H.V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Khali-dov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M., et al.: Dinov3. arXiv preprint arXiv:2508.10104 (2025)
24. Sitzmann, V., Martel, J., Bergman, A., Lindell, D., Wetzstein, G.: Implicit neural representations with periodic activation functions. Advances in neural information processing systems **33**, 7462–7473 (2020)

25. Vedaldi, A., Zisserman, A.: Vgg convolutional neural networks practical. Department of Engineering Science, University of Oxford **66**, 27 (2016)
26. Wu, J., Wang, Z., Hong, M., Ji, W., Fu, H., Xu, Y., Xu, M., Jin, Y.: Medical sam adapter: Adapting segment anything model for medical image segmentation. *Medical image analysis* **102**, 103547 (2025)
27. Yang, F., Lu, P.X., Deng, M., Wang, Y.X.J., Rajaraman, S., Xue, Z., Folio, L.R., Antani, S.K., Jaeger, S.: Annotations of lung abnormalities in the shenzhen chest x-ray dataset for computer-aided screening of pulmonary diseases. *Data* **7**(7), 95 (2022)
28. Yang, J., Shi, R., Wei, D., Liu, Z., Zhao, L., Ke, B., Pfister, H., Ni, B.: Medmnist v2- a large-scale lightweight benchmark for 2d and 3d biomedical image classification. *Scientific Data* **10**(1), 41 (2023)
29. Zech, J.R., Badgeley, M.A., Liu, M., Costa, A.B., Titano, J.J., Oermann, E.K.: Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. *PLoS medicine* **15**(11), e1002683 (2018)
30. Zhang, Q., Chen, M., Bukharin, A., He, P., Cheng, Y., Chen, W., Zhao, T.: Adaptive budget allocation for parameter-efficient fine-tuning. In: International Conference on Learning Representations. Openreview (2023)