Supplementary Material

A. Theoretical Analysis of Task Decorrelation

In this section, we provide a theoretical analysis showing that the proposed *frequency-switching* sine modulation decorrelates task-specific matrices derived from a shared low-rank base, whereas linear frequency scaling alone cannot. This analysis formally explains why the sine transformation is a necessary component for effective task specialization in our framework.

Correlation Definition. For two matrices $M_s, M_t \in \mathbb{R}^{m\times n}$, we define their correlation as the cosine similarity of their vectorized forms:

$$\mathrm{corr}(M_s, M_t) = \frac{\mathrm{vec}(M_s)^\top \mathrm{vec}(M_t)}{\|\mathrm{vec}(M_s)\|_2 \, \|\mathrm{vec}(M_t)\|_2}. \tag{16}$$

A value of 0 indicates decorrelation (orthogonality), while ± 1 indicates perfectly correlated matrices.

Proposition 1. *Let $M_{\mathrm{AWB}} = AWB^\top$ be the shared base matrix. Define task-specific kernels using frequency-dependent sine modulation:*

$$M_t = \sin(\omega_t M_{\mathrm{AWB}}), \qquad t = 1, \ldots, T, \tag{17}$$

where the sine is applied elementwise. Assume each entry of M_{AWB} is a zero-mean, finite-variance random variable with symmetric density. Then for any two distinct frequencies $\omega_s \neq \omega_t$, we have

$$\mathrm{corr}(M_s, M_t) \approx 0, \tag{18}$$

i.e., frequency-dependent sine mapping decorrelates the resulting task-specific matrices.

Proof. Let X be a scalar entry of M_{AWB}, and define $Y_s = \sin(\omega_s X)$ and $Y_t = \sin(\omega_t X)$. Using the identity $\sin(\alpha X)\sin(\beta X) = \frac{1}{2}\big(\cos((\alpha-\beta)X) - \cos((\alpha+\beta)X)\big)$, we obtain

$$\mathbb{E}[Y_s Y_t] = \tfrac{1}{2}\big(\mathbb{E}[\cos((\omega_s - \omega_t)X)] - \mathbb{E}[\cos((\omega_s + \omega_t)X)]\big).$$

Because X has a symmetric density and finite variance, the expectation $\mathbb{E}[\cos(kX)]$ goes to zero as $|k| \to \infty$. Therefore, whenever both $|\omega_s - \omega_t|$ and $|\omega_s + \omega_t|$ are large, the two cosine expectations are small, and thus

$$\mathbb{E}[Y_s Y_t] \approx 0.$$

Vectorizing M_t gives $\mathbf{m}_t = \mathrm{vec}(M_t)$, whose entries satisfy $M_t(i,j) \sim Y_t$, where $Y_t = \sin(\omega_t X) \in [-1, 1]$. By the law of large numbers,

$$\frac{\mathbf{m}_s^\top \mathbf{m}_t}{\|\mathbf{m}_s\|_2 \|\mathbf{m}_t\|_2} \to \frac{\mathbb{E}[Y_s Y_t]}{\sqrt{\mathbb{E}[Y_s^2]}\sqrt{\mathbb{E}[Y_t^2]}} \approx 0, \tag{19}$$

which proves $\mathrm{corr}(M_s, M_t) \approx 0$. □

Proposition 2. *Without sine transformation, frequency acts only as a scalar scaling and thus cannot decorrelate the AWB matrix. Define the linear-scaled task matrices*

$$\widetilde{M}_t = \omega_t M_{\mathrm{AWB}}. \tag{20}$$

Then for any tasks s, t we have

$$|\mathrm{corr}(\widetilde{M}_s, \widetilde{M}_t)| = 1, \tag{21}$$

i.e., the task-specific matrices are perfectly correlated and lie in the same one-dimensional subspace.

Proof. Vectorizing yields

$$\widetilde{\mathbf{m}}_t = \mathrm{vec}(\widetilde{M}_t) = \omega_t \, \mathrm{vec}(M_{\mathrm{AWB}}) = \omega_t \mathbf{m}. \tag{22}$$

For any s, t,

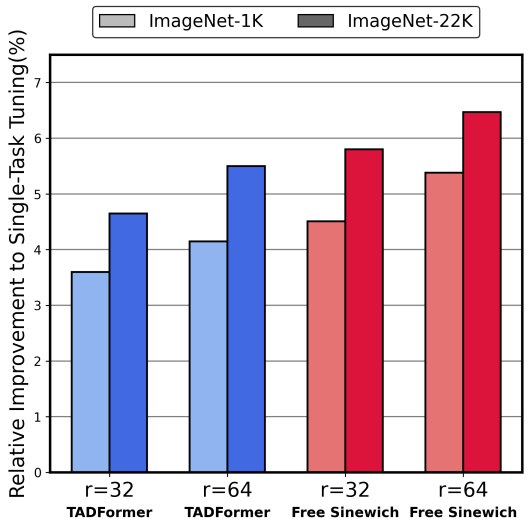
$$\mathrm{corr}(\widetilde{M}_s, \widetilde{M}_t) = \frac{(\omega_s \mathbf{m})^\top (\omega_t \mathbf{m})}{|\omega_s| \|\mathbf{m}\|_2 \, |\omega_t| \|\mathbf{m}\|_2} = \pm 1. \tag{23}$$

Thus all $\widetilde{M}_t$ are collinear scalings of the same matrix and share identical singular directions. Hence frequency without sine cannot create decorrelated or task-specific variations. □

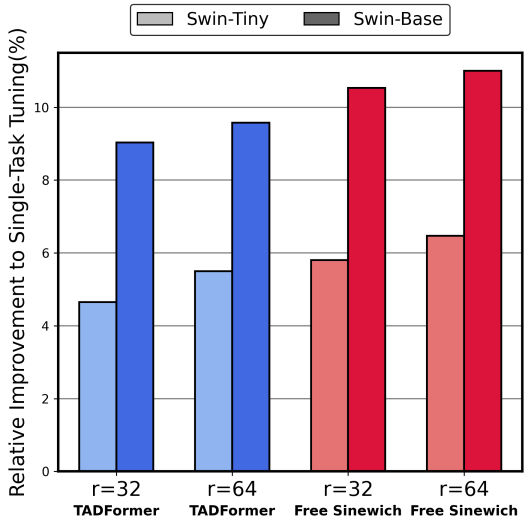
Implications. Together, Propositions 1 and 2 formally establish that the sine-based frequency modulation is the key mechanism enabling task-specific decorrelation from a shared low-rank base. Frequency alone, without the sine transformation, collapses all task-specific matrices into the same parameter subspace. This theoretical result directly supports the design of our frequency-switching Sine-AWB module and explains the empirical performance gains observed in our experiments.

B. Free Sinewich with Different Backbones and Pretraining Datasets

We evaluate the robustness of Free Sinewich under different backbone capacities and pretraining settings. Specifically, we examine the effect of (i) using a larger pretraining dataset and (ii) adopting a larger backbone architecture. The results are summarized in Fig. 5.



(a) Effect of pretraining dataset.



(b) Effect of backbone capacity.

Figure 5. **Effect of backbone pretraining and model capacity on Free Sinewich.** (a) Comparison between Swin-T pretrained on ImageNet-1K and ImageNet-22K. (b) Performance comparison between Swin-T and Swin-B backbones (both pretrained on ImageNet-22K).

As shown in Fig. 5a, pretraining on a larger dataset (ImageNet-22K) leads to a clear and consistent performance improvement over ImageNet-1K pretraining. This indicates that a stronger pretrained backbone provides a richer shared representation, which directly enhances the effectiveness of frequency switching. Since Free Sinewich reuses a single shared base matrix across tasks, the quality of the pretrained weights plays a crucial role in enabling effective task-specific modulation.

Fig. 5b further shows that increasing backbone capacity from Swin-T to Swin-B also improves performance. This confirms that Free Sinewich benefits from stronger representational capacity in the shared backbone. The relative gain from enlarging the backbone is bigger for Free Sinewich than for TADFormer. We attribute this to a effect: Free Sinewich already achieves strong performance by efficiently reusing shared parameters through frequency switching, this further scales up with a bigger model. These results demonstrate that Free Sinewich scales favorably with both better pretraining and larger backbones, while maintaining its core advantage of parameter-efficient task specialization through frequency-based modulation.

C. Gaussian Low-Pass Filter Hyperparameters

We analyze the sensitivity of Free Sinewich to the hyperparameters of the Gaussian low-pass filter applied after sine modulation. Specifically, we perform one-dimensional sweeps over the kernel size K and the standard deviation σ, and report the results in Table 6.

We first vary the kernel size while fixing $\sigma = 1.0$. As shown in the upper block of Table 6, the performance remains stable across different kernel sizes. Among the tested configurations, $K = 7$ achieves the best overall result, yielding the highest Δm (+5.39). Both smaller ($K = 5$) and larger ($K = 9$) kernels produce comparable performance, indicating that the method is not sensitive to the exact spatial extent of the filter.

We then fix the kernel size to $K = 7$ and sweep the standard deviation σ. The results show that the original setting $\sigma = 1.0$ consistently achieves the best trade-off across all tasks. When $\sigma = 1.5$, the performance remains competitive, demonstrating robustness to moderate over-smoothing. However, when $\sigma = 0.5$, we observe a noticeable drop in performance, with Δm decreasing from +5.39 to +4.90. We attribute this degradation to insufficient suppression of high-frequency artifacts introduced by the sine transformation. With a small σ, the Gaussian filter becomes too narrow to effectively smooth oscillatory noise, which adversely affects feature stability and task-specific modulation. These results show that Free Sinewich is robust to a wide range of Gaussian filter hyperparameters, while the default setting ($K = 7$, $\sigma = 1.0$) provides the most consistent and optimal performance.

D. Gradient Cosine Similarity Analysis

To better understand task interference, we measure the similarity between task gradients computed on the shared LoRA base matrices of the encoder. For a pair of tasks i and j, let g_i and g_j denote the flattened gradients of the shared parameters. Their cosine similarity is computed as

$$\mathrm{sim}(i, j) = \frac{g_i^\top g_j}{\|g_i\|_2 \, \|g_j\|_2}. \tag{24}$$

Table 6. **Ablation over low-pass filter hyperparameters.** Effect of kernel size K and standard deviation σ on performance.

Kernel Size k	Standard Deviation σ	SemSeg (mIoU ↑)	Human Parts (mIoU ↑)	Saliency (mIoU ↑)	Normals (rmse ↓)	Δm (%)
Sweep over kernel size K (fixed $\sigma = 1.0$)						
$K = 5$	$\sigma = 1.0$	71.20	61.37	66.16	16.15	+5.32
$K = 7$	$\sigma = 1.0$	**71.25**	61.38	**66.24**	16.14	**+5.39**
$K = 9$	$\sigma = 1.0$	71.06	**61.46**	66.07	**16.11**	+5.32
Sweep over σ (fixed $K = 7$)						
$K = 7$	$\sigma = 0.5$	70.91	61.26	65.58	16.17	+4.90
$K = 7$	$\sigma = 1.0$	**71.25**	**61.38**	66.24	16.14	**+5.39**
$K = 7$	$\sigma = 1.5$	71.05	61.31	**66.40**	**16.13**	+5.36

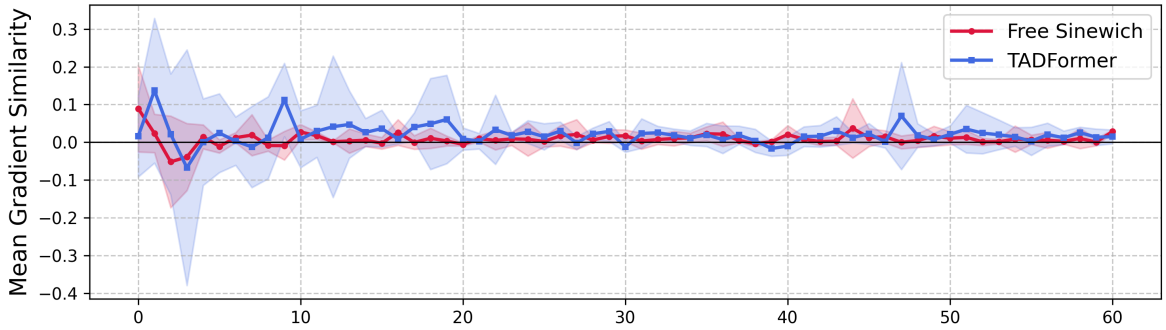


Figure 6. Pairwise gradient cosine similarity measured on the shared LoRA base matrix across training epochs (x-axis). Our method maintains near-orthogonal inter-task gradients with reduced variance and lower conflict rate compared to TADFormer.

During training, gradients from different tasks are collected at each iteration. The similarity values are accumulated across iterations and averaged within each epoch:

$$\bar{s} = \frac{1}{N} \sum_{k=1}^{N} \mathrm{sim}_k(i, j), \tag{25}$$

where N is the number of gradient samples in the epoch. We also report the variance of these values to quantify gradient stability.

Fig. 6 shows a comparison of gradient cosine similarity between TADFormer and Free Sinewich. Compared to TADFormer, our method exhibits lower variance and more stable, near-orthogonal inter-task gradients (similarity ≈ 0), indicating reduced interference despite shared parameters. Although we do not explicitly design a gradient optimization component, the proposed *frequency-switching* mechanism appears to implicitly induce task-specific subspaces within the shared base matrix. Combining our method with explicit gradient-based approaches [9, 41] remains a promising direction for future work.

Table 7. **Results on the Cityscapes dataset.** All methods use a Swin-Tiny backbone [26] pretrained on ImageNet-22K [7] with an HRNet decoder. MTL-Dec denotes training the decoder only, while MTL-Full denotes full fine-tuning. Δm denotes the average relative improvement over the single-task baseline.

Method	SemSeg (mIoU ↑)	Depth (rmse ↓)	Δm (%)	Trainable Param. (M)
Single Task	63.98	5.91	0	56.00
MTL - Dec	53.30	8.54	-30.59	0.97
MTL - Full	61.03	5.76	-1.03	28.49
DiTASK [27]	56.08	6.35	-9.89	2.58
MTLoRA [1]	60.88	5.84	-1.83	7.23
TADFormer [3]	62.76	5.39	+3.44	5.97
Free Sinewich	62.62	5.14	+5.45	5.92

E. Results on the Cityscapes Dataset

We further evaluate Free Sinewich on the Cityscapes [6] dataset to assess its generalization to large-scale urban scene understanding. Following the experimental protocols of TaskPrompter [39] and DiffusionMTL [40], we consider a two-task setting consisting of semantic segmentation and monocular depth estimation. All methods are trained and

evaluated under identical settings to ensure a fair comparison.

The quantitative results are summarized in Table 7. Among parameter-efficient multi-task learning methods, DiTASK and MTLoRA show limited improvements and remain below the single-task baseline. TADFormer improves upon these baselines, achieving a positive Δm of +3.44 with 5.97M trainable parameters. In contrast, Free Sinewich achieves the best overall performance, attaining a Δm of +5.45 with only 5.92M trainable parameters. These results indicate that Free Sinewich effectively mitigates task interference while maintaining strong parameter efficiency in large-scale outdoor scenes.