DWTSG: Parameter-Efficient Fine-Tuning of Large Pre-trained Models via Discrete Wavelet Transform and Subband Guidance

Chengwei Sun[1], Jiwei Wei[1,3*], Shiyuan He[1], Zeyu Ma [1], Yuyang Zhou [2], Ran Ran [1], Jie Zou [1], Yang Yang [1, 3]

[1]School of Computer Science and Engineering, University of Electronic Science and Technology of China
[2]School of Cyberspace Security, Hainan University
[3]Institute of Electronic and Information Engineering , University of Electronic Science and Technology of China
202311081544@std.uestc.edu.cn, mathematic6@gmail.com, shiyuanhe.david@gmail.com, cnzeyuma@163.com, yuyangzhouzhou@hainanu.edu.cn, ranran@std.uestc.edu.cn, jie.zou@uestc.edu.cn, yang.yang@uestc.edu.cn

Abstract

Fully fine-tuning large pre-trained models for each downstream task is impractical due to prohibitive memory, computation, and storage costs. Although parameter-efficient fine-tuning (PEFT) methods address this issue, leading methods like LoRA still exhibit linear scaling of trainable parameters with hidden size. Recent studies have explored PEFT in the frequency domain to reduce computational costs by employing fast Fourier transform and discrete cosine transform with sparse frequency selection. These methods rely on global frequency representations that lack spatial locality and disperse energy across the domain. As a result, sparse coefficient selection struggles to preserve fine-grained structural information and often introduces artifacts such as ringing near boundaries. To address these limitations, we propose DWTSG, a novel PEFT framework based on discrete wavelet transform (DWT) and subband guidance. DWTSG decomposes pre-trained weights into four wavelet subbands that jointly encode global context and local details. It fine-tunes only the most informative coefficients in each subband through an energy-based selection strategy that prioritizes coefficients based on their individual importance and interactions. Finally, inverse DWT reconstructs the updated weights, enabling efficient and precise adaptation. Extensive experiments on natural language understanding, commonsense reasoning, and image classification demonstrate that DWTSG outperforms existing PEFT methods, achieving superior performance and higher parameter efficiency.

Introduction

Large pre-trained models (LPMs) have achieved remarkable success across natural language processing and computer vision tasks. However, as these models continue to scale, fully fine-tuning them for each downstream task becomes increasingly impractical due to high memory, computation, and storage demands (Wolf et al. 2020; Brown et al. 2020; Ding et al. 2023). To mitigate these challenges, parameter-efficient fine-tuning (PEFT) methods have emerged, aiming to reduce overhead by updating only a small subset of parameters or incorporating lightweight modules (Hu et al. 2022; Zhang et al. 2023; Ren et al. 2024). Among these, LoRA (Hu et al. 2022) has become particularly popular by approximating weight updates with a pair of low-rank matrices, effectively reducing the number of trainable parameters. However, its trainable parameter count scales linearly with the hidden dimension, which can remain large in LPMs. This raises a critical question: Can we further reduce the number of trainable parameters by updating only the most structurally expressive components of the model weights, thereby achieving more efficient and precise task adaptation?

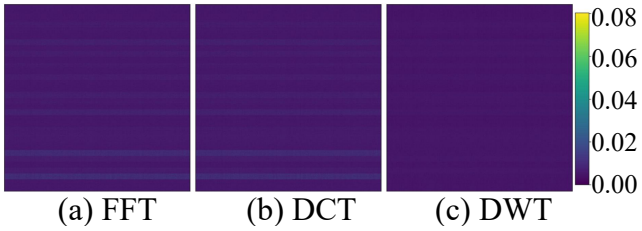

Figure 1: Residual heatmaps of reconstructed v_proj weight matrix from the first self-attention layer of LLaMA2-7B, where each reconstruction is based on the top 10% of energy coefficients preserved by (a) FFT, (b) DCT, and (c) DWT. FFT and DCT exhibit prominent horizontal stripe artifacts from global frequency truncation, whereas DWT yields a more uniform residual distribution, highlighting its superior preservation of localized structural details.

Recent efforts have explored this idea in the frequency domain. For instance, FourierFT (Gao et al. 2024) leverages fast Fourier transform (FFT) to learn sparse spectral coefficients that represent weight changes, while sDCTFT (Shen et al. 2024) employs the discrete cosine transform (DCT) and selects critical frequency components via predefined partitions. Although these methods effectively reduce the number of trainable parameters, their reliance on global frequency representations, where each coefficient influences the entire spatial domain, leads to three limitations. First, the lack of spatial locality makes it difficult to focus updates on structurally meaningful regions. Second, the energy in the global spectral domain is typically dispersed, making it

*Corresponding Author.

hard to identify and prioritize important coefficients. Third, retaining only a small subset of coefficients as trainable parameters often results in incomplete or distorted updates, as global frequency truncation fails to preserve fine-grained details and may introduce artifacts such as ringing near boundaries (Lu and Zhou 2019; Jiang et al. 2020; Sun et al. 2023).

In contrast, the discrete wavelet transform (DWT) addresses these limitations through its structural properties. First, DWT provides strong spatial locality through compactly supported basis functions, allowing weight updates to focus on spatially meaningful regions (Heil and Walnut 1989). Second, it performs multi-scale decomposition, yielding four subbands that separately encode low-frequency structures and high-frequency variations along different orientations. This supports localized modeling under sparse conditions while avoiding structured artifacts introduced by global truncation. Third, DWT offers strong energy compaction, concentrating informative content into a small number of coefficients and thereby facilitating efficient parameter selection (Dong et al. 2024). To validate these advantages, we analyze the reconstruction errors when only the top 10% of energy coefficients are retained after applying FFT, DCT, and DWT to the v_proj weight matrix from the first self-attention layer of LLaMA2-7B (Touvron et al. 2023). As shown in Figure 1, reconstructions based on FFT and DCT exhibit structured stripe artifacts and higher residual errors due to global frequency truncation, while the reconstructions based on DWT show significantly smaller and more uniformly distributed residual errors. This toy experiment indicates DWT's superior capability in preserving critical structural information through effective localization and compact representation, clearly highlighting its advantage over global frequency methods.

Motivated by these insights, we propose a novel PEFT framework based on the Discrete Wavelet Transform and Subband Guidance (DWTSG). DWTSG first applies a two-dimensional DWT to each pre-trained weight matrix, decomposing it into four structurally meaningful subbands. Within each subband, we adopt an energy-based selection strategy to identify and fine-tune only the most significant wavelet coefficients, guided by analyses of subband importance and interaction. After training, the updated coefficients are merged back into their respective subbands, and an inverse DWT (IDWT) is applied to reconstruct the adapted weight matrix in the spatial domain. The resulting update is then scaled and added to the original weights, enabling targeted adaptation with minimal overhead. Extensive experiments on natural language processing and computer vision tasks validate the efficiency of DWTSG. Furthermore, we analyze the impact of key design choices, including the selection of wavelet bases, strategies for subband parameter allocation, and the impact of the scaling factor.

The main contributions of this work are as follows:

- We introduce DWTSG, a novel PEFT framework that operates entirely in the wavelet domain by selectively updating the most structurally informative wavelet coefficients from each subband of pre-trained weight matrices.
- By analyzing the functional roles and interactions among wavelet subbands, DWTSG enables sparse yet structurally expressive parameter updates, allowing effective adaptation with minimal trainable parameters.
- Extensive experiments across natural language understanding, commonsense reasoning, and image classification show that DWTSG consistently achieves competitive or superior performance with significantly fewer trainable parameters than existing PEFT methods.

Related Works

PEFT methods aim to reduce the computational and GPU memory costs of adapting LPMs to downstream tasks. They can be broadly categorized into three major types: low-rank adaptation methods, weight decomposition-based methods, and spectral adaptation methods.

Low-rank adaptation methods use low-rank matrices to approximate weight updates. For example, LoRA (Hu et al. 2022) introduces two trainable low-rank matrices for each pre-trained model weight to approximate the desired updates of the original model (see Figure 2). Building on this, VeRA (Kopiczko, Blankevoort, and Asano 2024) improves efficiency by sharing one pair of low-rank matrices across all layers and using layer-specific scaling vectors, reducing the parameter count significantly. MELoRA (Ren et al. 2024) stacks multiple low-rank LoRA modules in parallel, with each mini-LoRA learning distinct subspaces of the hidden representations, achieving better performance with fewer parameters.

The second category consists of **weight decomposition-based methods**, which decompose pre-trained weights into trainable and frozen components. PiSSA (Meng, Wang, and Zhang 2024) applies singular value decomposition to the pre-trained weight matrix, splitting it into principal components and a residual matrix. The principal components containing the most significant singular values initialize trainable low-rank matrices. The residual matrix, which contains smaller singular values, remains frozen during fine-tuning. This selective updating of the most important subspace enables faster convergence. In contrast, MiLoRA (Wang et al. 2025) focuses on updating the minor singular components while keeping the principal components fixed. This is based on the idea that task-specific information can be captured in the less dominant subspaces of the model.

The third category is **spectral adaptation methods**, which fine-tune models in the frequency domain by updating a subset of spectral coefficients. FourierFT (Gao et al. 2024) learns a small number of frequency coefficients via FFT and reconstructs the weight matrix through inverse transformation. sDCTFT (Shen et al. 2024) follows a similar approach using DCT, which provides better energy compaction and decorrelation. It projects low-rank weight updates into the cosine domain, partitions them across spectral levels, and selects the most critical frequency components for fine-tuning.

Method

In this section, we first transform pre-trained weight matrices into the wavelet domain, producing four subbands. We then apply a selective parameter allocation strategy that

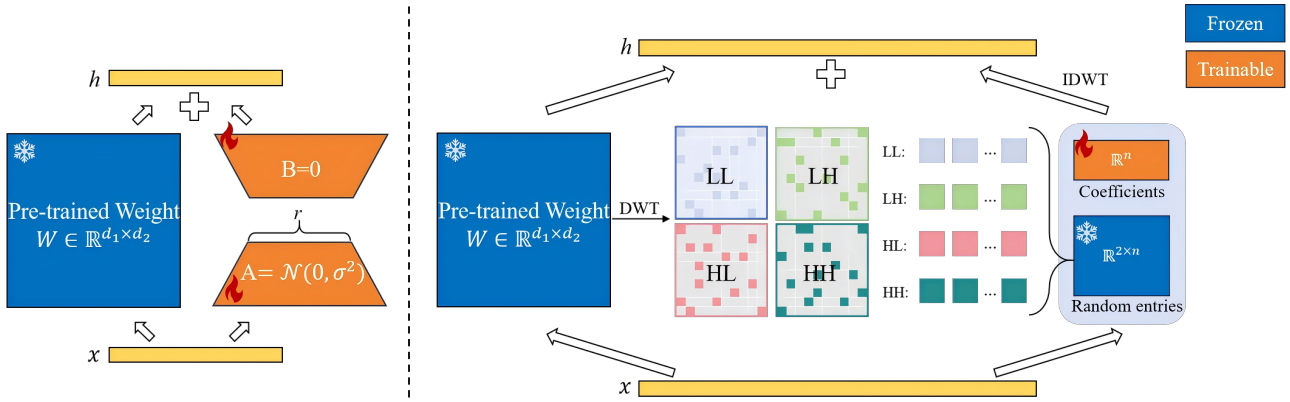

Figure 2: Comparison between the LoRA structure (left) and the proposed DWTSG method (right). LoRA models weight updates using two low-rank matrices, where B is initialized to zero and A is drawn from a Gaussian distribution. In contrast, DWTSG applies a two-dimensional DWT to the pre-trained weight matrix, yielding four subbands: LL, LH, HL, and HH. A set of trainable coefficients is then selected from each subband based on energy criteria. After training, the updated coefficients are reconstructed via IDWT and merged with the original weights to form the final updated weights.

focuses training on a sparse subset of high-magnitude coefficients within each subband, substantially reducing the number of trainable parameters. To guide this process, we systematically analyze the functional roles and interactions among subbands, enabling optimal allocation of trainable parameters. Finally, the IDWT is used to reconstruct the updated weights in the spatial domain, allowing for precise and efficient model adaptation. An overview of the proposed DWTSG framework is presented in Figure 2.

Weight Matrix Transformation via DWT

Given a pre-trained weight W, where $W[i, j]$ denotes the original weight values, we apply a two-dimensional DWT to obtain its frequency domain representation. Taking the Symlets wavelet basis (sym4) as an example, which is known for its orthogonality and approximate symmetry (Daubechies 1992), the associated filter bank includes a decomposition low-pass filter $g[\ell]$, a decomposition high-pass filter $h[\ell]$, and their reconstruction filters $\tilde{g}[\ell]$ and $\tilde{h}[\ell]$. The high-pass filter is derived from the low-pass filter via reflection and sign alternation: $h[\ell] = (-1)^{\ell} \cdot g[L-1-\ell]$, where $\ell \in [0, L-1]$ is the filter coefficient index and $L = 8$ for sym4.

The two-dimensional DWT is applied by sequentially performing one-dimensional wavelet transforms along the rows and columns of W, resulting in four subbands through two stages of convolution and downsampling. In the first stage, each row of W is convolved with $g[\ell]$ and $h[\ell]$, followed by downsampling, producing intermediate matrices A and D that capture the low and high frequency components:

$$\begin{aligned} A[i,m] &= \sum_{\ell=0}^{7} g[\ell] \cdot W[i, 2m+\ell-4], \\ D[i,m] &= \sum_{\ell=0}^{7} h[\ell] \cdot W[i, 2m+\ell-4], \end{aligned} \tag{1}$$

where m is the column index after downsampling.

In the second stage, column-wise convolution and downsampling are applied to A and D to produce final subbands:

$$\begin{aligned} \mathrm{LL}[n,m] &= \sum_{\ell=0}^{7} g[\ell] \cdot A[2n+\ell-4, m], \\ \mathrm{LH}[n,m] &= \sum_{\ell=0}^{7} h[\ell] \cdot A[2n+\ell-4, m], \\ \mathrm{HL}[n,m] &= \sum_{\ell=0}^{7} g[\ell] \cdot D[2n+\ell-4, m], \\ \mathrm{HH}[n,m] &= \sum_{\ell=0}^{7} h[\ell] \cdot D[2n+\ell-4, m], \end{aligned} \tag{2}$$

where n is the row index after downsampling.

Selection Strategy for Trainable Parameters

Upon obtaining the wavelet subbands, we develop a coefficient selection framework tailored to the structure of each subband to determine which coefficients should be trainable. For each subband matrix $B \in \mathbb{R}^{h \times w}$, we define a binary selection function $\mathcal{M}(B, p)$ that identifies the top-p elements with the largest absolute values:

$$\mathcal{M}(B,p) = M, \quad M_{i,j} = \begin{cases} 1, & \text{if } (i,j) \in \mathcal{I}_p(B), \\ 0, & \text{otherwise}, \end{cases} \tag{3}$$

where $\sum_{i,j} M_{i,j} = p$ and $\mathcal{I}_p(B)$ denotes the index set of the p largest magnitude entries in B.

During training, only the coefficients corresponding to $M_{i,j} = 1$ are updated, while the rest remain frozen. These selected values are concatenated into a flat parameter vector $\theta \in \mathbb{R}^p$, enabling sparse and efficient updates directly in the frequency domain. With this selection mechanism in place, allocating the limited trainable parameters across different subbands is the following key challenge. Addressing this requires understanding the individual roles and interactions of each subband in relation to model performance.

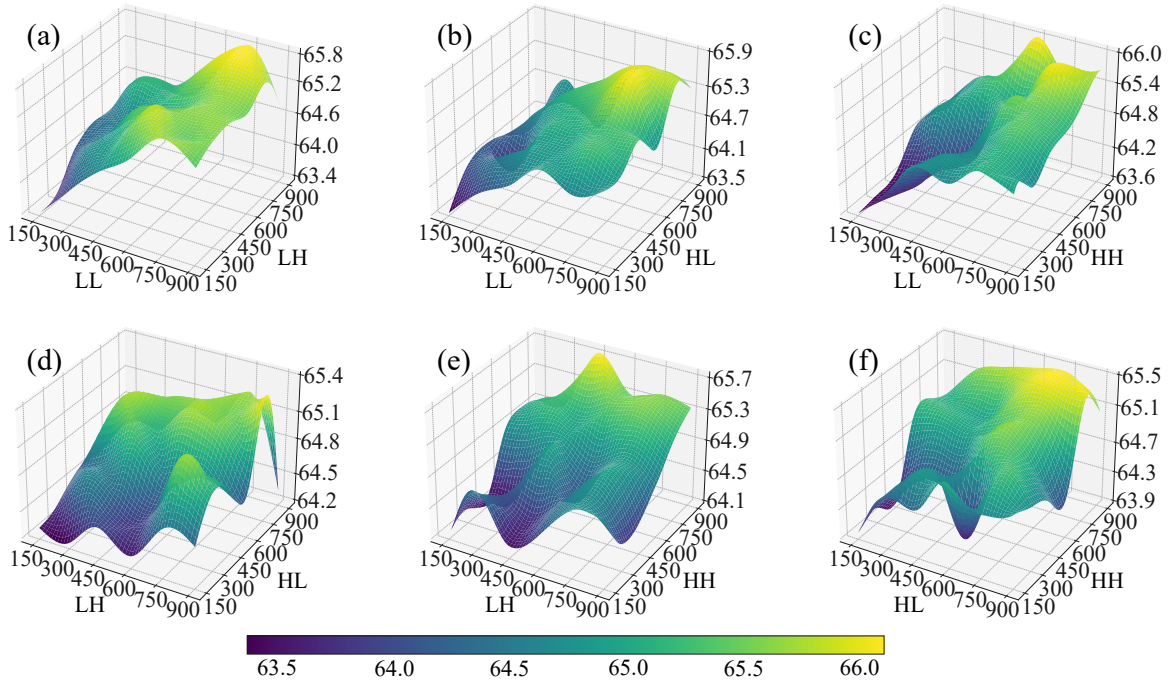

Figure 3: 3D surface plots showing the impact of joint parameter allocation across wavelet subband pairs in PEFT using RoBERTa-large on the CoLA dataset. Each subplot shows performance as a function of two subband parameter counts, with the remaining subbands averaged over all configurations. Subplots (a)–(f) correspond to the six pairwise combinations: (a) LL–LH, (b) LL–HL, (c) LL–HH, (d) LH–HL, (e) LH–HH, and (f) HL–HH. The x-axes and y-axes represent the number of trainable parameters allocated to the two subbands, and the z-axis shows the MCC.

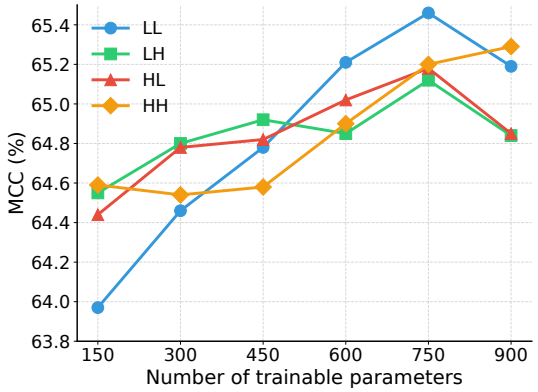

Figure 4: Effect of trainable parameter allocation on individual wavelet subbands using RoBERTa-large on the CoLA dataset. The x-axis denotes the number of trainable parameters, and the y-axis reports the corresponding MCC.

Functional Roles of Wavelet Subbands

We systematically evaluate the individual contributions of wavelet subbands by conducting an extensive grid search on the RoBERTa-large model fine-tuned on the CoLA dataset. The Matthew's correlation coefficient (MCC) is used as the evaluation metric. For each subband, the number of trainable parameters is independently varied within the candidate set $\{150, 300, 450, 600, 750, 900\}$, and the results are averaged over all parameter configurations of the remaining subbands.

As shown in Figure 4, the LL subband exerts the most substantial influence on model performance, with MCC consistently increasing from 64.0% to 65.5% as more parameters are allocated. This underscores its critical role in capturing low-frequency global structures. The HL subband also contributes stable gains, with MCC rising from 64.4% to 65.1%, suggesting its importance in modeling horizontal structural features. The HH subband reaches a competitive MCC of 65.3% when fully parameterized, indicating its effectiveness in capturing fine-grained high-frequency details. In contrast, the LH subband exhibits relatively weaker gains, with performance improvements primarily occurring in mid-to-high parameter ranges, likely due to its sensitivity to localized vertical features.

Synergistic Effects Among Wavelet Subbands

To gain deeper insight into the interactions between wavelet subbands, we perform a comprehensive grid search over all pairwise combinations of subband parameter allocations and visualize the results using 3D surface plots, as shown in Figure 3. Figures 3(a)–(c) demonstrate that increasing the LL subband's trainable parameters consistently improves performance across all pairings, reinforcing its dominant role in representing global semantics. In contrast, the LH subband yields limited marginal benefits, as seen in Figures 3(a), (d), and (e), and offers only modest gains even when com-

Method	Params.	SST-2	MRPC	CoLA	QNLI	RTE	STS-B	MNLI	QQP	Avg.
FF	356M	96.3	**90.9**	68.0	94.7	86.6	92.4	90.2	**92.2**	**88.9**
LoRA	0.8M	96.2	90.2	68.2	**94.8**	85.2	92.3	**90.6**	91.6	88.6
AdaLoRA	0.8M	95.9	90.6	67.9	94.2	86.4	**92.7**	**90.6**	91.6	88.7
VeRA	0.061M	96.1	**90.9**	68.0	94.4	85.9	91.7	90.2	90.3	88.4
FourierFT	0.048M	96.0	**90.9**	67.1	94.4	**87.4**	91.9	88.9	89.2	88.2
DWTSG	0.144M	**96.7**	90.7	**69.5**	94.3	86.6	92.1	90.0	90.5	88.8

Table 1: Performance comparison of fine-tuning methods on the GLUE benchmark using RoBERTa-large. The best result for each dataset is highlighted in bold. Higher values indicate better performance across all evaluation metrics.

bined with HL or HH. Meanwhile, both HL and HH subbands demonstrate stable and complementary enhancement effects, as illustrated in Figures 3(b), (c), and (f). Specifically, the HL–HH combination yields the most pronounced synergistic gain, indicating their effectiveness in modeling directional and fine-grained structural patterns.

Overall, LL serves as the core semantic component and should receive the largest share of trainable parameters. HL and HH enhance structural expressiveness through directional and high-frequency cues and are best configured with balanced parameter budgets. In contrast, LH offers limited auxiliary support and can be allocated minimal parameters. These insights guide effective allocation strategies for DWTSG under varying budget constraints. For instance, in the Experiments section, we adopt a default allocation of 30% to LL, 15% to LH, 30% to HL, and 25% to HH, which consistently yields strong results. While broadly effective, task-specific tuning may further enhance performance.

Reconstruction and Integration via IDWT

After training, the updated coefficient vector θ is mapped back to its original positions within the corresponding wavelet subbands. We then employ the IDWT to reconstruct the matrix $W_{\text{re}} \in \mathbb{R}^{d_1 \times d_2}$. The reconstruction is performed in two steps. First, row-wise synthesis is performed:

$$\begin{aligned} A[i,m] &= \sum_{\ell=0}^{7} \tilde{g}[\ell] \cdot \text{LL}[\lfloor (i-\ell+4)/2 \rfloor, m] \\ &\quad + \sum_{\ell=0}^{7} \tilde{h}[\ell] \cdot \text{LH}[\lfloor (i-\ell+4)/2 \rfloor, m], \\ D[i,m] &= \sum_{\ell=0}^{7} \tilde{g}[\ell] \cdot \text{HL}[\lfloor (i-\ell+4)/2 \rfloor, m] \\ &\quad + \sum_{\ell=0}^{7} \tilde{h}[\ell] \cdot \text{HH}[\lfloor (i-\ell+4)/2 \rfloor, m], \end{aligned} \tag{4}$$

then followed by column-wise synthesis:

$$\begin{aligned} W_{\text{re}}[i,j] &= \sum_{\ell=0}^{7} \tilde{g}[\ell] \cdot A[i, \lfloor (j-\ell+4)/2 \rfloor] \\ &\quad + \sum_{\ell=0}^{7} \tilde{h}[\ell] \cdot D[i, \lfloor (j-\ell+4)/2 \rfloor]. \end{aligned} \tag{5}$$

The frequency domain weight update is then computed as the difference between the reconstructed weight matrix W_{re} and the original weight matrix W:

$$\Delta W = W_{\text{re}} - W. \tag{6}$$

Finally, the update is scaled by a tunable hyperparameter α to control the adaptation intensity:

$$W' = W + \alpha \times \Delta W. \tag{7}$$

Experiments

Models and Datasets

We evaluate DWTSG across three domains.

For **natural language understanding**, we use the GLUE benchmark (Wang et al. 2019), which consists of eight tasks: SST-2 (Socher et al. 2013), MRPC (Dolan and Brockett 2005), CoLA (Warstadt, Singh, and Bowman 2019), QNLI (Rajpurkar et al. 2016), RTE (Giampiccolo et al. 2007), STS-B (Cer et al. 2017), MNLI (McCoy, Min, and Linzen 2019), and QQP (Yang et al. 2021). These cover a range of tasks, including sentiment classification, paraphrase detection, sentence similarity, and natural language inference. We fine-tune the RoBERTa-large model using various PEFT methods (Liu et al. 2019). Evaluation metrics include MCC for CoLA, Pearson's correlation coefficient for STS-B, and classification accuracy for the remaining tasks.

For **commonsense reasoning**, we fine-tune LLaMA2-7B on the Commonsense170K dataset (Hu et al. 2023), a collection of eight benchmarks: BoolQ (Clark et al. 2019), PIQA (Bisk et al. 2020), SIQA (Sap et al. 2019), HellaSwag (Zellers et al. 2019), WinoGrande (Sakaguchi et al. 2020), ARC-e (Clark et al. 2018), ARC-c (Clark et al. 2018), and OBQA (Mihaylov et al. 2018). Following the LLM-Adapters setup, we treat each as a multiple-choice task and select the best checkpoint based on validation loss. We report accuracy using the LLM-Adapters implementation.

For **image classification**, we evaluate on eight datasets: OxfordPets (Parkhi et al. 2012), StanfordCars (Krause et al. 2013), CIFAR-10 (Krizhevsky, Hinton et al. 2009), DTD (Cimpoi et al. 2014), EuroSAT (Helber et al. 2019), FGVC (Maji et al. 2013), RESISC45 (Cheng, Han, and Lu 2017), and CIFAR-100 (Krizhevsky, Hinton et al. 2009). Both ViT-base and ViT-large models are fine-tuned, using ImageNet-21k pre-trained weights (Dosovitskiy et al. 2021; Ridnik et al. 2021; Wei et al. 2024a,b).

Further details regarding the datasets and hyperparameters are provided in the supplementary materials.

Method	Params.	BoolQ	PIQA	SIQA	HellaSwag	WinoGrande	ARC-e	ARC-c	OBQA	Avg.
LoRA	56.6M	69.8	79.9	79.5	83.6	82.6	79.8	64.7	**81.0**	77.6
PiSSA	83.8M	67.6	78.1	78.4	76.6	78.0	75.8	60.2	75.6	73.8
MiLoRA	56.6M	67.6	83.8	**80.1**	**88.2**	82.0	82.8	68.8	80.6	79.2
DWTSG	25.6M	**71.3**	**84.2**	**80.1**	87.6	**83.0**	**83.8**	**69.5**	80.8	**80.0**

Table 2: Performance comparison of LLaMA2-7B fine-tuned with different PEFT methods on eight commonsense reasoning benchmarks. The best score for each dataset is highlighted in bold.

	Method	Params.	OxfordPets	StanfordCars	CIFAR10	DTD	EuroSAT	FGVC	RESISC45	CIFAR100	Avg.
Base	LP	-	90.3	25.8	96.4	69.8	88.7	17.4	74.2	84.3	68.4
	FF	85.8M	93.1	**79.8**	98.9	77.7	**99.1**	**54.8**	**96.1**	**92.4**	**86.5**
	LoRA	581K	93.2	45.4	98.8	75.0	98.4	25.2	92.7	92.0	77.6
	FourierFT	72K	93.2	46.1	98.6	75.1	98.3	27.5	92.0	91.2	77.7
	DWTSG	72K	**96.7**	58.8	**99.2**	**79.4**	98.8	39.7	94.5	91.7	82.4
Large	LP	-	91.1	37.9	97.8	73.3	92.6	24.6	82.0	84.3	73.0
	FF	303.3M	94.4	**88.9**	99.2	81.8	**99.0**	**68.3**	**96.4**	93.6	**90.2**
	LoRA	1.57M	94.8	73.3	99.1	81.8	98.6	42.3	94.7	**94.9**	84.9
	FourierFT	144K	94.5	69.6	99.1	80.8	98.7	39.9	93.9	93.3	83.7
	DWTSG	144K	**98.4**	82.5	**99.5**	**84.0**	98.9	63.5	95.9	94.0	89.6

Table 3: Performance comparison of fine-tuning methods on image classification tasks using ViT-base and ViT-large models. Accuracy is reported after ten training epochs on each dataset. The best result for each dataset is highlighted in bold.

Baselines

We compare DWTSG with several representative PEFT methods across different tasks. For natural language understanding, the baselines include Full Fine-tuning (FF) (Huang et al. 2022), LoRA (Hu et al. 2022), AdaLoRA (Zhang et al. 2023), VeRA (Kopiczko, Blankevoort, and Asano 2024), and FourierFT (Gao et al. 2024). For commonsense reasoning, we compare with LoRA, PiSSA (Meng, Wang, and Zhang 2024), and MiLoRA (Wang et al. 2025). For image classification, we consider linear probing (LP) (Gao et al. 2024), FF, LoRA, and FourierFT.

Comparison Results

Natural Language Understanding. Table 1 shows results on the GLUE benchmark using RoBERTa-large. FF achieves the highest average score but requires updating all 356M parameters, making it computationally expensive. Among PEFT methods, DWTSG is particularly notable, achieving an impressive average score of 88.8% with only 0.144M trainable parameters, nearly matching FF and surpassing other PEFT baselines. DWTSG excels on CoLA, demonstrating its ability to capture fine-grained linguistic features, and performs competitively on SST-2 and MRPC, showing robustness across different task types. LoRA and AdaLoRA deliver strong overall results with slightly larger parameter counts of 0.8M, averaging scores of 88.6% and 88.7%, respectively. AdaLoRA excels in tasks like STS-B and RTE due to its adaptive parameter allocation strategy. VeRA and FourierFT are notable for their compact size, with VeRA matching FF on MRPC and FourierFT achieving the best score on RTE. However, their overall average performance is slightly lower, indicating a trade-off between extreme compression and generalization. In a word, DWTSG provides an outstanding balance between model efficiency and accuracy, making it particularly advantageous for resource-constrained environments.

Commonsense Reasoning. As shown in Table 2, we evaluate several PEFT methods on LLaMA2-7B across eight commonsense reasoning benchmarks. DWTSG uses only 25.6M trainable parameters, significantly fewer than LoRA, PiSSA, and MiLoRA, achieving substantial parameter reductions. Despite this, DWTSG achieves the highest average accuracy among all methods. It shows notable performance gains on BoolQ, PIQA, WinoGrande, and ARC datasets, demonstrating its ability to handle complex reasoning tasks. On SIQA and HellaSwag, DWTSG performs similarly to MiLoRA, slightly trailing on HellaSwag. This suggests that MiLoRA's ensemble approach may be advantageous for nuanced contextual reasoning. DWTSG effectively balances efficiency and performance, showing great potential as an efficient PEFT solution for LPMs.

Image Classification. Table 3 reports performance on eight image classification datasets using ViT-base and ViT-large. DWTSG consistently achieves high accuracy with significantly fewer trainable parameters than other methods. For ViT-base, our method reaches the highest average accuracy of 82.4%, outperforming LoRA and FourierFT. It also surpasses FF on five of eight datasets, using only 72K trainable parameters. For ViT-large, our method achieves an average accuracy of 89.6%, close to FF's 90.2%, but with only 144K trainable parameters. This demonstrates that DWTSG is highly effective and efficient for computer vision tasks.

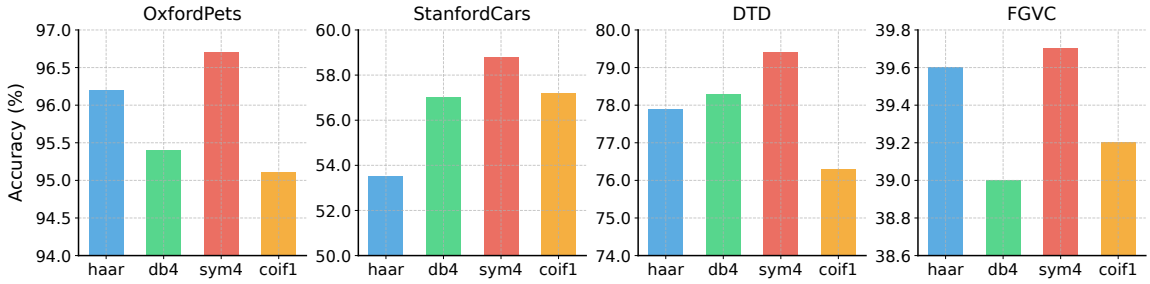

Figure 5: Classification accuracy of four wavelet bases on four image classification datasets under the DWTSG framework.

Dataset	Random	Energy	Ours
OxfordPets	95.6	94.6	**96.7**
StanfordCars	56.0	55.7	**58.8**
DTD	77.0	79.2	**79.4**
FGVC	36.9	38.5	**39.7**

Table 4: Comparison of ViT-base accuracy across four datasets using three subband parameter selection strategies.

Dataset	0.1	0.3	0.5	0.7	0.9	1.1
OxfordPets	93.1	96.2	94.0	93.2	**96.7**	93.5
StanfordCars	46.5	52.8	56.8	57.3	**58.8**	55.1
DTD	74.8	**79.4**	77.9	77.0	75.3	75.2
FGVC	34.6	**39.7**	35.9	27.6	22.0	5.1

Table 5: Classification accuracy of ViT-base on four image classification datasets with different scaling factor values α.

Ablation Study

We analyze key design choices for DWTSG, including wavelet basis selection, subband parameter allocation strategies, and the scaling factor α. Experiments are performed on four image classification datasets using the ViT-base model.

Wavelet Basis Selection. To assess the impact of wavelet basis selection on DWTSG, we fine-tune ViT-base using four representative wavelet families: Haar, Daubechies (db4), Symlets (sym4), and Coiflets (coif1). As illustrated in Figure 5, Symlets consistently achieves the highest accuracy across all datasets, reaching 96.7% on OxfordPets, 58.8% on StanfordCars, 79.4% on DTD, and 39.7% on FGVC, surpassing the other bases by up to 5.3 percentage points. While Haar is computationally efficient, its limited smoothness restricts performance. Coiflets offer minor improvements over Haar but are still less effective than Daubechies and Symlets. These findings underscore the importance of selecting wavelets with favorable structural properties such as smoothness and near symmetry, as exemplified by the strong performance of Symlets in PEFT tasks.

Subband Parameter Allocation Strategies. We compare three strategies for selecting trainable parameters per weight matrix in the wavelet domain after two-dimensional DWT. The first, Random, distributes parameters uniformly across subbands, assuming equal importance. The second, Energy, allocates parameters based on each subband's squared ℓ_2 norm, selecting the largest-magnitude coefficients within each quota. The third, our proposed method, incorporates subband roles and interactions: LL functions as the semantic backbone, HL and HH serve as complementary detail enhancers, and LH plays a more auxiliary role. Based on this, we define heuristic allocation ranges and select top coefficients within each subband accordingly. As shown in Table 4, our strategy consistently outperforms the other two across all datasets. While Energy improves over Random by leveraging subband energy, it underperforms on fine-grained tasks like FGVC. These results underscore the importance of accounting for both subbands' individual roles and interactions when allocating parameters.

Effect of Scaling Factor α. Table 5 summarizes the influence of the scaling factor α in the weight update formula $W' = W + \alpha \times \Delta W$ across different image classification tasks. For datasets with abundant training data or coarse-grained categories such as OxfordPets and StanfordCars, higher values of α like 0.9 yield peak performance of 96.7% and 58.8%, respectively, highlighting benefits from more substantial weight adjustments. Conversely, fine-grained or resource-limited tasks such as DTD and FGVC achieve optimal accuracy at moderate scaling factors, notably at $\alpha = 0.3$, with accuracies of 79.4% and 39.7%. Larger α values tend to reduce performance, underscoring the importance of task-specific tuning to preserve critical feature distinctions.

Conclusion

This work introduced DWTSG, a novel PEFT framework that operated in the wavelet domain. By selectively updating the most structurally informative components of pre-trained weights, DWTSG significantly reduced the number of trainable parameters while maintaining or surpassing the performance of existing methods. Extensive experiments across natural language understanding, commonsense reasoning, and image classification tasks demonstrated that DWTSG successfully balanced efficiency and effectiveness. Beyond empirical results, it provided valuable insights into critical design choices such as wavelet basis selection and subband-wise parameter allocation, highlighting the potential of wavelet transforms for efficient adaptation of LPMs and inspiring future research in this direction.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grants 62306067, 62220106008), the Sichuan Science and Technology Program (Grants 2024NSFSC1463, 2025ZNSFSC0479), the Guangdong Basic and Applied Basic Research Foundation (2025A1515010108), and the Sichuan Province Innovative Talent Funding Project for Postdoctoral Fellows (BX202405).

References

Bisk, Y.; Zellers, R.; Gao, J.; Choi, Y.; et al. 2020. PIQA: Reasoning about Physical Commonsense in Natural Language. In *AAAI Conference on Artificial Intelligence*, volume 34, 7432–7439.

Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J. D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, 1877–1901.

Cer, D.; Diab, M.; Agirre, E.; Lopez-Gazpio, I.; and Specia, L. 2017. Semeval-2017 task 1: Semantic textual similarity-multilingual and cross-lingual focused evaluation. *arXiv:1708.00055*.

Cheng, G.; Han, J.; and Lu, X. 2017. Remote Sensing Image Scene Classification: Benchmark and State of the Art. *Proc. IEEE*, 105(10): 1865–1883.

Cimpoi, M.; Maji, S.; Kokkinos, I.; Mohamed, S.; and Vedaldi, A. 2014. Describing textures in the wild. In *IEEE Conference on Computer Vision and Pattern Recognition*, 3606–3613.

Clark, C.; Lee, K.; Chang, M.; Kwiatkowski, T.; Collins, M.; and Toutanova, K. 2019. BoolQ: Exploring the Surprising Difficulty of Natural Yes/No Questions. In *Association for Computational Linguistics*, 2924–2936.

Clark, P.; Cowhey, I.; Etzioni, O.; Khot, T.; Sabharwal, A.; Schoenick, C.; and Tafjord, O. 2018. Think you have Solved Question Answering? Try ARC, the AI2 Reasoning Challenge. *arXiv:1803.05457*.

Daubechies, I. 1992. *Ten Lectures on Wavelets*. SIAM. ISBN 978-0-89871-274-2.

Ding, N.; Qin, Y.; Yang, G.; Wei, F.; Yang, Z.; Su, Y.; Hu, S.; Chen, Y.; Chan, C.-M.; Chen, W.; et al. 2023. Parameter-efficient fine-tuning of large-scale pre-trained language models. *Nat. Mach. Intell.*, 5(3): 220–235.

Dolan, B.; and Brockett, C. 2005. Automatically Constructing a Corpus of Sentential Paraphrases. In *Proceedings of the Third International Workshop on Paraphrasing*.

Dong, C.; Ma, H.; Li, Z.; Li, L.; and Liu, D. 2024. Temporal Wavelet Transform-Based Low-Complexity Perceptual Quality Enhancement of Compressed Video. *IEEE Trans. Circuits Syst. Video Technol.*, 34(5): 4040–4053.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; Uszkoreit, J.; and Houlsby, N. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *International Conference on Learning Representations*.

Gao, Z.; Wang, Q.; Chen, A.; Liu, Z.; Wu, B.; Chen, L.; and Li, J. 2024. Parameter-Efficient Fine-Tuning with Discrete Fourier Transform. In *International Conference on Machine Learning*.

Giampiccolo, D.; Magnini, B.; Dagan, I.; and Dolan, W. B. 2007. The Third PASCAL Recognizing Textual Entailment Challenge. In *Association for Computational Linguistics*, 1–9.

Heil, C. E.; and Walnut, D. F. 1989. Continuous and discrete wavelet transforms. *SIAM REV.*, 31(4): 628–666.

Helber, P.; Bischke, B.; Dengel, A.; and Borth, D. 2019. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, 12(7): 2217–2226.

Hu, E. J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; and Chen, W. 2022. LoRA: Low-Rank Adaptation of Large Language Models. In *International Conference on Learning Representations*.

Hu, Z.; Wang, L.; Lan, Y.; Xu, W.; Lim, E.; Bing, L.; Xu, X.; Poria, S.; and Lee, R. K. 2023. LLM-Adapters: An Adapter Family for Parameter-Efficient Fine-Tuning of Large Language Models. In *Conference on Empirical Methods in Natural Language Processing*, 5254–5276.

Huang, S.; Xu, D.; Yen, I. E.; Wang, Y.; Chang, S.-E.; Li, B.; Chen, S.; Xie, M.; Rajasekaran, S.; Liu, H.; et al. 2022. Sparse progressive distillation: Resolving overfitting under pretrain-and-finetune paradigm. In *Annual Meeting of the Association for Computational Linguistics*, 190–200.

Jiang, T.; Ng, M. K.; Zhao, X.; and Huang, T. 2020. Framelet Representation of Tensor Nuclear Norm for Third-Order Tensor Completion. *IEEE Trans. Image Process.*, 29: 7233–7244.

Kopiczko, D. J.; Blankevoort, T.; and Asano, Y. M. 2024. VeRA: Vector-based Random Matrix Adaptation. In *International Conference on Learning Representations*.

Krause, J.; Stark, M.; Deng, J.; and Fei-Fei, L. 2013. 3D Object Representations for Fine-Grained Categorization. In *IEEE International Conference on Computer Vision*, 554–561.

Krizhevsky, A.; Hinton, G.; et al. 2009. Learning multiple layers of features from tiny images. *Master's thesis, University of Tront*.

Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; and Stoyanov, V. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv:1907.11692*.

Lu, C.; and Zhou, P. 2019. Exact recovery of tensor robust principal component analysis under linear transforms. *arXiv:1907.08288*.

Maji, S.; Rahtu, E.; Kannala, J.; Blaschko, M.; and Vedaldi, A. 2013. Fine-Grained Visual Classification of Aircraft. *arXiv:1306.5151*.

McCoy, R. T.; Min, J.; and Linzen, T. 2019. BERTs of a feather do not generalize together: Large variability in generalization across models with similar test set performance. *arXiv:1911.02969*.

Meng, F.; Wang, Z.; and Zhang, M. 2024. PiSSA: Principal Singular Values and Singular Vectors Adaptation of Large Language Models. In *Advances in Neural Information Processing Systems*.

Mihaylov, T.; Clark, P.; Khot, T.; and Sabharwal, A. 2018. Can a Suit of Armor Conduct Electricity? A New Dataset for Open Book Question Answering. In *Conference on Empirical Methods in Natural Language Processing*, 2381–2391.

Parkhi, O. M.; Vedaldi, A.; Zisserman, A.; and Jawahar, C. 2012. Cats and dogs. In *IEEE Conference on Computer Vision and Pattern Recognition*, 3498–3505.

Rajpurkar, P.; Zhang, J.; Lopyrev, K.; and Liang, P. 2016. SQuAD: 100000+ Questions for Machine Comprehension of Text. In *Conference on Empirical Methods in Natural Language Processing*, 2383–2392.

Ren, P.; Shi, C.; Wu, S.; Zhang, M.; Ren, Z.; de Rijke, M.; Chen, Z.; and Pei, J. 2024. MELoRA: Mini-Ensemble Low-Rank Adapters for Parameter-Efficient Fine-Tuning. In *Annual Meeting of the Association for Computational Linguistics*, 3052–3064.

Ridnik, T.; Baruch, E. B.; Noy, A.; and Zelnik, L. 2021. ImageNet-21K Pretraining for the Masses. In *Advances in Neural Information Processing Systems*.

Sakaguchi, K.; Bras, R. L.; Bhagavatula, C.; and Choi, Y. 2020. Winogrande: An adversarial winograd schema challenge at scale. In *AAAI Conference on Artificial Intelligence*, volume 64, 99–106.

Sap, M.; Rashkin, H.; Chen, D.; LeBras, R.; and Choi, Y. 2019. SocialIQA: Commonsense Reasoning about Social Interactions. *arXiv:1904.09728*.

Shen, Y.; Bi, Q.; Huang, J.-H.; Zhu, H.; and Pathania, A. 2024. Parameter-Efficient Fine-Tuning via Selective Discrete Cosine Transform. *arXiv:2410.09103*.

Socher, R.; Perelygin, A.; Wu, J.; Chuang, J.; Manning, C. D.; Ng, A. Y.; and Potts, C. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Conference on Empirical Methods in Natural Language Processing*, 1631–1642.

Sun, C.; Huang, T.; Xu, T.; and Deng, L. 2023. NF-3DLogTNN: An effective hyperspectral and multispectral image fusion method based on nonlocal low-fibered-rank regularization. *Appl Math Model*, 118: 780–797.

Touvron, H.; Martin, L.; Stone, K.; Albert, P.; Almahairi, A.; Babaei, Y.; Bashlykov, N.; Batra, S.; Bhargava, P.; Bhosale, S.; et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv:2307.09288*.

Wang, A.; Singh, A.; Michael, J.; Hill, F.; Levy, O.; and Bowman, S. R. 2019. GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding. In *International Conference on Learning Representations*, 353–355.

Wang, H.; Li, Y.; Wang, S.; Chen, G.; and Chen, Y. 2025. MiLoRA: Harnessing Minor Singular Components for Parameter-Efficient LLM Finetuning. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics*, 4823–4836.

Warstadt, A.; Singh, A.; and Bowman, S. R. 2019. Neural network acceptability judgments. *Trans. Assoc. Comput. Linguist.*, 7: 625–641.

Wei, J.; Pan, C.; He, S.; Wang, G.; Yang, Y.; and Shen, H. T. 2024a. Towards Robust Person Re-Identification by Adversarial Training With Dynamic Attack Strategy. *IEEE Trans. Multim.*, 26: 10367–10380.

Wei, J.; Yang, Y.; Guan, X.; Xu, X.; Wang, G.; and Shen, H. T. 2024b. Runge-Kutta Guided Feature Augmentation for Few-Sample Learning. *IEEE Trans. Multim.*, 26: 7349–7358.

Wolf, T.; Debut, L.; Sanh, V.; Chaumond, J.; Delangue, C.; Moi, A.; Cistac, P.; Rault, T.; Louf, R.; Funtowicz, M.; Davison, J.; Shleifer, S.; von Platen, P.; Ma, C.; Jernite, Y.; Plu, J.; Xu, C.; Scao, T. L.; Gugger, S.; Drame, M.; Lhoest, Q.; and Rush, A. M. 2020. Transformers: State-of-the-art natural language processing. In *Conference on Empirical Methods in Natural Language Processing*, 38–45.

Yang, W.; Li, L.; Zhang, Z.; Ren, X.; Sun, X.; and He, B. 2021. Be Careful about Poisoned Word Embeddings: Exploring the Vulnerability of the Embedding Layers in NLP Models. In *The North American Chapter of the Association for Computational Linguistics*, 2048–2058.

Zellers, R.; Holtzman, A.; Bisk, Y.; Farhadi, A.; and Choi, Y. 2019. HellaSwag: Can a Machine Really Finish Your Sentence? In *Association for Computational Linguistics*, 4791–4800.

Zhang, Q.; Chen, M.; Bukharin, A.; Karampatziakis, N.; He, P.; Cheng, Y.; Chen, W.; and Zhao, T. 2023. Adaptive Budget Allocation for Parameter-Efficient Fine-Tuning. In *International Conference on Learning Representations*.