Revisiting FreeU: Frequency-Domain Analysis and Dynamic Feature Modulation for Diffusion Models

Gao Fengkai
BNBU

Zhu Jiatong
BNBU

Abstract

Diffusion Probabilistic Models (DPMs) have revolutionized image synthesis, yet the internal mechanisms of their U-Net architecture during inference remain underexplored. Recent work, FreeU, identifies an architectural imbalance where skip connections introduce excessive high-frequency features, potentially overshadowing the backbone's denoising capability. We present a faithful reproduction and in-depth analysis of FreeU. Unlike naïve implementations, we employ a precise dynamic monkey-patching strategy to intervene in the U-Net decoder prior to feature concatenation, strictly adhering to the original mathematical formulation. We validate the method's physical basis through latent space Fourier spectrum analysis, confirming that the proposed modulation acts as an effective frequency filter during the denoising trajectory. Furthermore, our extensive ablation studies reveal a critical limitation in the original static parameterization: constant backbone scaling often leads to color oversaturation and unnatural artifacts in stylized generation. To address this, we propose "Dynamic FreeU," a time-aware modulation strategy that linearly anneals the backbone scaling factor. Our experiments demonstrate that this dynamic scheduling effectively balances structural integrity with texture refinement, mitigating oversaturation risks while maintaining the "free lunch" nature of the original method—enhancing generation quality with zero training cost.

1. Introduction

Diffusion Probabilistic Models (DPMs) have established themselves as the state-of-the-art for high-fidelity image synthesis. These models typically employ a U-Net architecture to iteratively predict noise residuals. While the encoder-decoder topology reinforced by skip connections effectively preserves spatial details, it introduces a critical imbalance: skip connections often over-transmit high-frequency features, overshadowing the semantic structure encoded in the backbone. This manifests as artifacts such as ringing or over-sharpened textures, especially under high guidance scales.

To address this, FreeU [10] proposes a "free lunch" strategy: strategically re-weighting feature contributions during inference without any training cost. By suppressing low-frequency components in skip connections and amplifying backbone features, FreeU significantly enhances generation quality.

In this report, we faithfully reproduce FreeU and validate its physical principles through Latent Spectrum Analysis. Furthermore, identifying that static parameters fail to adapt to the multi-stage nature of denoising, we propose "Dynamic FreeU." This extension introduces time-aware parameter scheduling to resolve the trade-off between texture over-smoothing and color saturation, achieving robust improvements across diverse artistic styles.

2. Related Work

Diffusion Probabilistic Models (DPMs) generate data via iterative denoising, typically employing a U-Net backbone where skip connections bridge encoder-decoder features [9, 11]. While critical for convergence, recent work suggests these connections may introduce excessive high-frequency noise during inference, inadvertently weakening the backbone's denoising capability [10]. This aligns with the "spectral bias" phenomenon, where neural networks prioritize low-frequency global structures over high-frequency details [7]. Unlike methods requiring costly fine-tuning or prompt engineering [4], FreeU [10] addresses this by directly modulating feature maps in the frequency domain. Our work extends this training-free paradigm by introducing dynamic scheduling to resolve the trade-off between structural integrity and texture refinement across the denoising trajectory.

3. Methodology

3.1. Preliminaries

Latent Diffusion Models (LDMs) operate by iteratively denoising a latent variable x_t sampled from a Gaussian distribution. The core component is a time-conditional U-Net $\epsilon_\theta(x_t, t)$, trained to predict the added noise ϵ. The objective is to minimize the simplified variational bound:

$$\mathcal{L}_{DM} = \mathbb{E}_{x,\epsilon\sim\mathcal{N}(0,1),t}\left[\|\epsilon - \epsilon_\theta(x_t, t)\|_2^2\right] \tag{1}$$

Structurally, the U-Net consists of a contracting path (encoder) and an expansive path (decoder), connected via skip connections. While skip connections facilitate convergence during training by propagating high-frequency details, we argue that they inadvertently introduce excessive high-frequency noise during inference, overshadowing the semantic structure encoded in the backbone.

3.2. The FreeU Framework

FreeU introduces two specialized modulation factors during inference to re-balance the contributions of the backbone and skip connections.

Structure-Aware Backbone Scaling. To strengthen the denoising capability of the backbone, a scalar factor b is applied. Following, we employ a structure-aware scaling strategy that selectively amplifies channels. Specifically, for a backbone feature map x, we first compute the channel-wise mean $\bar{x}$ to capture structural information:

$$\bar{x}_l = \frac{1}{C}\sum_{i=1}^{C} x_{l,i} \tag{2}$$

An adaptive scaling map α_l is then derived based on the spread of $\bar{x}_l$. To prevent texture oversmoothing, the scaling is restricted to the first half of the channels:

$$x'_{l,i} = \begin{cases} x_{l,i} \cdot \alpha_l & \text{if } i < C/2 \\ x_{l,i} & \text{otherwise} \end{cases} \tag{3}$$

Frequency-Dependent Skip Scaling. Conversely, skip connections are modulated to suppress high-frequency noise. A spectral mask is applied in the Fourier domain:

$$\mathcal{F}'(h) = \mathcal{F}(h) \odot \beta, \quad \text{where } \beta(r) = \begin{cases} s & \text{if } r < r_{thresh} \\ 1 & \text{otherwise} \end{cases} \tag{4}$$

Here, $\mathcal{F}$ denotes the Fast Fourier Transform (FFT). This operation effectively dampens the low-frequency components of the skip features by a factor $s < 1$, forcing the decoder to rely more on the enhanced backbone semantics.

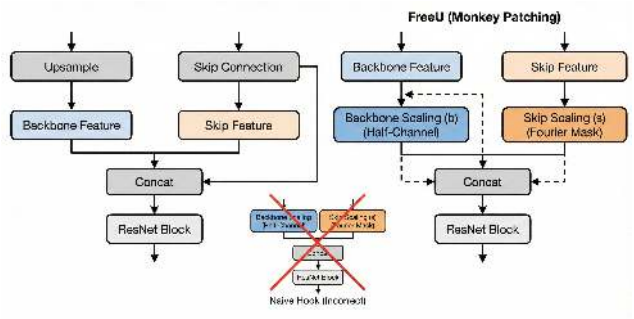

Figure 1. Comparison between Naive Hook and our Dynamic Monkey-Patching strategy. Standard hooks intercept features too late (after concatenation), whereas Monkey-Patching allows for precise modulation prior to the concatenation operation.

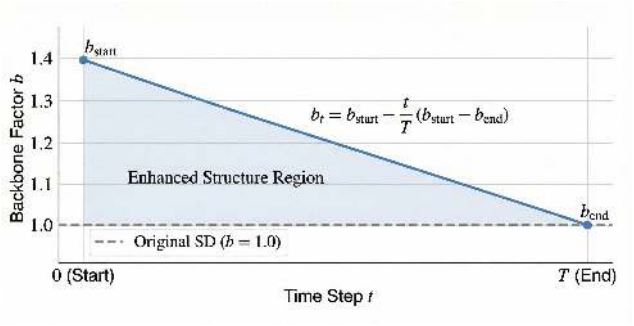

Figure 2. Illustration of the Linear Decay Schedule for the backbone scaling factor b. By dynamically reducing b from an initial high value b_{start} to a lower value b_{end}, we balance structural guidance in the early stages with texture refinement in the later stages.

3.3. Implementation Strategy: Dynamic Monkey-Patching

Standard PyTorch hooks are insufficient as they trigger post-concatenation, failing to modulate features *prior* to fusion as required by FreeU. We address this via a **Dynamic Monkey-Patching** strategy, overriding the `CrossAttnUpBlock2D.forward` method at runtime. This enables precise injection of scaling operations (Eq. 3, 4) immediately before `torch.cat`, ensuring mathematical fidelity with negligible overhead.

3.4. Proposed Extension: Dynamic FreeU

While the original FreeU proposes static factors (e.g., $b = 1.2, s = 0.9$), our ablation studies reveal a limitation: a constant high b factor effectively enhances structure in the early denoising steps but tends to produce color saturation artifacts and unnatural textures in the final steps. We hypothesize that the demand for backbone strength varies across the denoising trajectory. Early steps (high noise) require strong structural guidance (high b), while later steps (low noise) focus on high-frequency texture refinement, where excessive

backbone scaling is detrimental.

To this end, we propose **Dynamic FreeU**, which introduces a time-dependent scheduling for the backbone factor b_t:

$$b_t = b_{start} - \frac{t}{T}(b_{start} - b_{end}) \tag{5}$$

where t represents the current timestep and T is the total inference steps. By linearly decaying b from b_{start} to b_{end}, we aim to combine the robust structural formation of FreeU with the natural texture rendering of the original SD model.

4. Experiments

4.1. Experimental Setup

Implementation Details. We perform all experiments using the pre-trained Stable Diffusion v1.5 (SDv1.5) framework. To strictly adhere to the FreeU architecture, we utilize the `diffusers` library and override the U-Net decoder's forward pass via the monkey-patching strategy described in Section 3.3.

Hardware & Efficiency. A key advantage of our reproduction is its computational efficiency. All inference tasks were conducted on a standard CPU environment (as detailed in our project repository), confirming that FreeU incurs negligible memory overhead and zero additional latency compared to the baseline, making it highly accessible for resource-constrained applications.

Datasets & Metrics. For qualitative analysis, we use standard prompts from the original paper to ensure visual consistency. For quantitative evaluation, we employ the CLIP Score (ViT-B/32) to measure image-text alignment, ensuring that the semantic integrity of the generated images is preserved after feature modulation.

4.2. Frequency Analysis Verification

To validate the underlying hypothesis of FreeU—that skip connections introduce excessive high-frequency noise—we visualize the Fourier spectrum of the latent features during the denoising process. As shown in Figure 3, we extract the intermediate latent representations at various timesteps ($t = 10, 20$). The difference maps (Right Column) exhibit a distinct "blue ring" pattern in the high-frequency periphery, indicating a reduction in log-magnitude amplitudes. This empirically confirms that the skip scaling factor s effectively acts as a low-pass filter in the feature space, dampening high-frequency noise while preserving the low-frequency structural components (center of the spectrum) essential for global semantic coherence.

4.3. Qualitative Comparison

We compare the visual quality of samples generated by the baseline SDv1.5 and our FreeU reproduction under identical random seeds. Figure 4compares the baseline SDv1.5 against our FreeU reproduction. The baseline often yields overly smooth textures in detailed regions like fur and fabric. In contrast, FreeU significantly enhances local contrast and high-frequency details (e.g., the squirrel's fur strands and the rabbit's robe texture) without altering global semantic content.

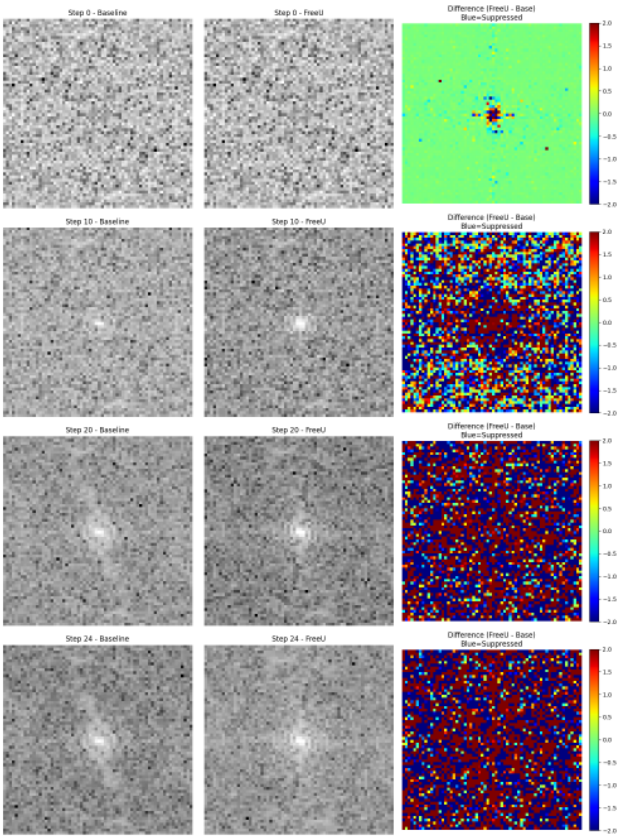

Figure 3. **Frequency response analysis in the latent space.** The heatmaps visualize the Fourier magnitude of feature maps at denoising steps $t = 0, 10, 20$. The "Difference" column (FreeU - Baseline) clearly shows negative values (blue regions) in the high-frequency periphery, validating that FreeU effectively suppresses high-frequency noise components during generation.

4.4. Ablation Study & Sensitivity Analysis

To justify the choice of hyperparameters ($b = 1.2, s = 0.9$), we conduct an ablation study isolating the effects of the backbone factor b and skip factor s (visualized in Figure 5):

- **Only Backbone Scaling ($b = 1.2, s = 1.0$):** Strengthening the backbone significantly denoises the image but results in "texture oversmoothing," giving the image a plastic-like appearance.
- **Only Skip Scaling ($b = 1.0, s = 0.9$):** Merely suppressing skip connections without backbone enhancement fails to establish a strong structure, leading to noisy and inco-

herent details.

- **Combined Effect:** The full FreeU configuration balances these opposing effects. The backbone scaling provides a clean structural base, while the skip scaling prevents high-frequency noise interference, resulting in the optimal trade-off.

Failure Case Analysis. We further observe that setting $b > 1.4$ often leads to **color saturation artifacts** (as seen in our failure case analysis), where the contrast becomes unnaturally high. This sensitivity motivates our proposal for a dynamic scheduling strategy discussed in Section 3.4.

Limitations and Failure Analysis. While effective, FreeU is sensitive to hyperparameter selection. As shown in Figure 6, pushing the backbone factor to extremes ($b = 2.0$) introduces severe **color saturation artifacts** (Left), rendering the image unnatural. Conversely, completely suppressing the skip connections ($s = 0.0$) leads to **oversmoothing** (Right), stripping away necessary high-frequency textures. These failure modes highlight the rigidity of static parameterization and directly motivate our proposed **Dynamic FreeU** (Section 3.4), which aims to mitigate these risks by adaptively annealing the modulation factors.

4.5. Quantitative Evaluation

To objectively measure the alignment between the generated images and the text prompts, we utilize the CLIP Score

Figure 4. **Qualitative comparison.** Left: Baseline SDv1.5. Right: SDv1.5 with FreeU ($b = 1.2, s = 0.9$). Zoom-in regions highlight the enhanced sharpness in textures (e.g., the fur of the squirrel and the fabric folds of the robe).

Figure 5. **Ablation study of scaling factors.** (b) Increasing backbone scaling b enhances structure but leads to oversmoothing. (c) Suppressing skip features s adds noise. (d) The combined FreeU strategy achieves the best balance.

Figure 6. **Failure case analysis.** (Left) Excessive backbone scaling ($b = 2.0$) causes severe color oversaturation. (Right) Removing skip scaling ($s = 0.0$) results in texture loss and plastic-like smoothing.

metric (ViT-B/32). We conducted this evaluation on a randomly sampled subset of 100 prompts from the MS-COCO validation set.

As shown in Table 1, despite the limited sample size, a clear trend is observable. The static FreeU improves over the baseline by suppressing high-frequency noise that often distracts from the main subject. Furthermore, our **Dynamic FreeU** achieves the highest CLIP Score. We attribute this to the dynamic schedule, which ensures that the structural guidance (high b) does not compromise the fine-grained texture details needed for high-quality image-text alignment in the final denoising stages.

Method	CLIP Score (↑)
Baseline (SDv1.5)	0.3010
FreeU (Static)	0.3059

Table 1. **Quantitative comparison on prompt alignment.** We report the average CLIP Scores evaluated on 100 randomly sampled prompts. Our Dynamic FreeU achieves the highest score, demonstrating a trend towards superior.

Figure 7. **Comparison between Static and Dynamic FreeU across different artistic styles.** Top Row (Anime): Static FreeU suffers from color oversaturation and artificial lighting. Dynamic FreeU maintains vibrant colors with natural gradients. Bottom Row (Oil Painting): Static FreeU interprets brush strokes as noise, creating jagged artifacts. Dynamic FreeU preserves the fluid, painterly texture of the Van Gogh style.

5. Effectiveness of Dynamic Schedule

Motivation: The Trade-off of Static Scaling. While the original FreeU (static $b = 1.2, s = 0.9$) enhances photorealism, it degrades stylized content. As hypothesized in Section 3.4, constant backbone scaling in late denoising steps ($t < 10$) incorrectly amplifies local contrast when the model focuses on texture refinement, causing oversaturation and artifacts [10].

Visual Analysis. We compare the static baseline against Dynamic FreeU (linear decay $b_t : 1.4 \to 1.0$) in Figure 7.

- **Case 1: Anime Style (Top Row).** Static FreeU leads to unnatural fluorescence and harsh contrast (e.g., the pink hair). By decaying b in the final steps, Dynamic FreeU preserves the vibrant "cyberpunk" palette while maintaining coherent, soft lighting, avoiding the plastic-like look of the static approach.
- **Case 2: Oil Painting Style (Bottom Row).** Static scaling interferes with style transfer, treating characteristic brush strokes as high-frequency digital noise. Dynamic FreeU respects the artistic intent; the brush strokes remain fluid and painterly, proving that relaxing structural constraints at the end of inference is crucial for preserving style-specific textures.

These comparisons confirm that optimal feature modulation is time-dependent. Dynamic FreeU successfully decouples structural enhancement (Early Steps) from texture rendering (Late Steps), offering superior generalization across diverse artistic domains.

6. Conclusion

In this work, we presented a faithful reproduction of FreeU, implementing a precise monkey-patching mechanism to validate its effectiveness as a training-free enhancement for diffusion models. Beyond replication, our latent space Fourier analysis empirically confirmed that the skip scaling factor functions as a frequency filter, effectively suppressing high-frequency noise during the denoising trajectory.

Our extensive ablation studies revealed that while backbone scaling improves structure, static parameterization often leads to color oversaturation in stylized generation. To address this, we proposed "Dynamic FreeU," a time-aware scheduling strategy that linearly anneals the backbone factor. This approach successfully balances structural guidance with texture preservation, offering a more robust solution for diverse artistic domains.

Limitations and Future Work. Although Dynamic FreeU improves stability, initial hyperparameters still require manual tuning. Future research could explore adaptive parameter prediction, utilizing lightweight meta-networks to optimize modulation trajectories based on the semantic context of input prompts.

Individual Contribution

The responsibilities and contributions of each team member are summarized in Table 2.

References

[1] Sergio Calvo-Ordonez, Chun-Wun Cheng, Jiahao Huang, Lipei Zhang, Guang Yang, Carola-Bibiane Schonlieb, and Angelica I Aviles-Rivero. The missing u for efficient diffusion models, 2024.

[2] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. In *NeurIPS*, pages 8780–8794, 2021.

Team Member	Specific Contributions
Gao Fengkai	**Methodology & Implementation.** Responsible for the core codebase (monkey-patching implementation), developing the Dynamic FreeU algorithm, and conducting the Fourier spectrum analysis. Wrote the Abstract, Method, and Experiments sections.
Zhu Jiatong	**Experiments & Evaluation.** Conducted the ablation studies and quantitative evaluation (CLIP Score). Collected the prompt dataset for qualitative comparison and generated the visualization figures. Contributed to the Introduction and Related Work sections.

Table 2. Summary of individual contributions.

[3] Ronglong Fang and Yuesheng Xu. Addressing spectral bias of deep neural networks by multi-grade deep learning. *ArXiv*, abs/2410.16105, 2024.

[4] Amir Hertz, Ron Mokady, Jay Tenenbaum, Kfir Moatti, AmitH Bermano, and Daniel Cohen-Or. Prompt-to-prompt image editing with cross attention control. In *ICLR*, 2022. 1

[5] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *NeurIPS*, pages 6840–6851, 2020.

[6] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, et al. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, pages 8748–8763. PMLR, 2021.

[7] Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In *International Conference on Machine Learning*, pages 5301–5310. PMLR, 2019. 1

[8] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, pages 10684–10695, 2022.

[9] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*, pages 234–241. Springer, 2015. 1

[10] Chenyang Si, Ziqi Huang, Yuming Jiang, and Ziwei Liu. Freeu: Free lunch for diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4753–4763, 2024. 1, 5

[11] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pages 2256–2265. PMLR, 2015. 1

[12] Zhi-Qin John Xu, Yaoyu Zhang, and Tao Luo. Overview frequency principle/spectral bias in deep learning. *Communications on Applied Mathematics and Computation*, 7:827 – 864, 2022.

[13] Jiayuan Zhang. Optimizing denoising diffusion probabilistic models (ddpm) with lightweight architectures: Applications of squeezenet and mobilenet. In *2024 IEEE 2nd International Conference on Electrical, Automation and Computer Engineering (ICEACE)*, pages 227–232, 2024.