# Image Cross-Domain Generation and Adaptation: A Frequency-Domain Decoupling Perspective

Qiwang Gao

{2130929201@shu.edu.cn}

School of Computer Engineering and Science, Shanghai University, Shanghai, China

**Abstract.** This paper takes a systematic approach to summarize the literature on the frequency-domain view of image generation and cross-domain transfer. Mechanisms analysis, Fourier and wavelet transforms are some of the underlying mechanisms that are analyzed to aid the construction of a systematic classification framework. The framework has a set of 5 basic operation mechanisms and three transfer levels: input space, feature space, and model/latent space. The frequency domain approaches are successful in addressing the bottlenecks of the spatial domain approaches in managing the long-range dependencies, attenuating the artifacts, and aligning cross-domain distributions. This effectiveness is attributable to inherent statistical decoupling characteristics as well as global modeling capabilities. The history of technological development makes evident a consistent progression of the external data alignment technology to feature-level calibration, and finally to latent-space intervention and reconstruction at the architecture level operator.

**Keywords:** Frequency Domain Analysis; Image Synthesis; Domain Adaptation; Feature Decoupling; Fourier Transform

## 1 Introduction

Traditional deep learning approaches to the synthesis of images first paid attention to quality problems at the pixel level. As an example, traditionally, structural constraints were suggested in classic architectures such as Pix2Pix. This method constrains the spatial interdependency between the pixels, which is a widely adopted approach to reduce the image blurring artifacts that are typically experienced in per-pixel regression problems [1].

Nevertheless, the technical complexity of these spatial-domain-based approaches is greatly augmented when applying them to Unsupervised Domain Adaptation (UDA) tasks. The common UDA models can require the specialized training of auxiliary neural networks and the development of complex auxiliary loss functions. The task is to make the model learning domain-invariant representations of source and target domains. This method requires difficult and inherently unstable adversarial training processes. As a result, it is often accompanied by serious convergence challenges and substantial computational overhead [2].

ICIAAI 2026, May 22-24, Singapore, Singapore

DOI 10.4108/eai.22-5-2026.2365354

The essence of such complexity can be seen in the intricate spatial correlations of image components within the space domain (pixel space). Semantic content and stylistic features are very closely coupled at pixel-level representations at a pixel-wise level. Such a binding makes conventional alignment schemes ineffective in the face of drastic domain disparities, including those between simulation and reality. It, therefore, becomes a hard task to do style transfer without structural violations, a case that often causes loss of semantic consistency.

To deal with these problems, the frequency-domain perspective presents a non-parametric alternative with physical interpretability. It has been shown that images exhibit inherent statistical decoupling properties. Amplitude spectrum largely captures global style and statistical properties and the phase spectrum retains structural semantics and geometric shapes. Based on this property, a set of parameter-free techniques can be utilized to match domain gaps significantly using Fourier Domain Adaptation (FDA), enabling domain alignment through a simple spectrum swap without retraining [3]. In addition, the performance of adversarial perturbation in the frequency domain has been introduced to enable the models to recognize and boost sensitivity to key frequency bands. This hence results in a more robust transfer without the usage of source domain data [4]. In addition, the structural consistency constraint of the low-frequency branch is made possible by the multi-scale properties of the Discrete Wavelet Transform (DWT), which allows one to restrict the structural consistency. In the meantime, the high-frequency channel restores textures, which is an effect that successfully eliminates aliasing artifacts of generative processing inherent in traditional convolutions [5].

The paradigm avoids the instability of the adversarial training and can be trained with a strong semantic preservation at reduced computational complexity. Moreover, there are great benefits in a frequency-domain perspective, in terms of enhancing data privacy and system robustness. The amplitude spectrum permits domain alignment without losing core semantic privacy, which can be considered a natural desensitization mechanism of privacy-preserving Source-Free Domain Adaptation (SFDA). The aspect of frequency-domain intervention is a critical direction, creating a critical path to model transfer and image generation in the context of Sustainable AI by optimizing the trade-off between computational efficiency and high-fidelity performance.

The proposed paper will attempt to review systematically the technological development of image generation and cross-domain transfer from a frequency-domain viewpoint. To begin with, this study will draw a classification regime of Fourier and wavelet transforms based on the physical properties of Fourier and wavelet transforms and will include five fundamental operation mechanisms and three transfer levels. The research then examines how the global spectrum alignment can be converted to multi-scale architecture reconstruction using the new developments of diffusion and large-scale visual models. It makes comparisons in a horizontal way of the common algorithms in terms of quality of generation, efficiency, and security robustness. Lastly, the paper presents a perspective on a future path such as phase-aware modeling, adaptive frequency band modulation, as well as offers theoretical descriptions and technical considerations to construct effective, resilient, and interpretable Generative AI.

# 2 Categorization of Technical Mechanisms for Frequency-Domain Generation and Adaptation

The current focus of research work in the frequency domain can be roughly divided into two dimensions, including operation mechanism and transfer levels.

### 2.1 Cross-Domain Image Synthesis and Adaptation Based on Operational Mechanisms

The existing studies intervene in the frequency spectrum using five fundamental mechanisms.

First, Spectral Mixing has the impact of bringing style similarity through the exchange of low-frequency components between source and target domains. An example of such a study is Fourier Domain Adaptation (FDA). This approach proves to be quite useful in the transfer of semantic segmentation across domains.

Secondly, Frequency Randomization perturbs semantically irrelevant frequency bands. As an example, Frequency-Spatial Domain Randomization (FSDR) applies randomized noise to frequency ranges that are sensitive to the change in the domain. This strategy is helpful in improving information generalization by models under out-of-distribution scenarios [6].

At the same time, Spectral Supervision results in the use of a spectral penalty term in the training loss, which regularizes high-frequency energy. This process is useful in reducing generative artifacts, e.g., the checkerboard pattern [7].

Moreover, spectral decomposition uses analysis based on multi-scale decomposition such as the Discrete Wavelet Transform (DWT). An example is WaveFill, which decomposes an image into several sub-bands to deal with global and local structures respectively. Building on this, Hu et al. introduced the Wavelet Diffusion Neural Operator (WDNO), which integrates wavelet-based spectral decomposition into the diffusion process. By modeling the generative trajectory in the wavelet domain, this approach produces high-resolution synthetic samples with superior structural fidelity. Such advancements demonstrate the efficacy of spectral decomposition in generating high-quality datasets for robust AI training. [8].

Lastly, Spectral Filtering or Attention mechanisms use the Convolution Theorem. This method converts the multifaceted modeling of the global spatial relationships to the multiplication of the elements by the frequency domain. As a result, it attains a global receptive field and a high amount of computational overhead is substantially reduced. Adequate representatives are the Adaptive Fourier Neural Operator (AFNO), which substitutes self-attention with processing high-resolution images effectively [9]. On the same note, Global Filter Networks (GFNet) are successful in modeling long-range dependencies within the frequency domain in an efficient manner [10]. Moreover, frequency-domain filters such as FreeU improve image details produced by them by dynamically modulating the frequency ratio of features through inference [11].

### 2.2 Cross-Domain Image Synthesis and Adaptation Based on Transfer Levels

It is possible to classify frequency-domain interventions into 3 levels based on the functional stage, to which deep neural networks belong.

The first is the Input Space Level, which entails data-level preprocessing. At this stage, the image spectrum is adjusted through the Fourier transformation to minimize domain discrepancy.

Then comes the Feature Space level, where the interventions aid in the control of latent representations to attain a refined calibration of deep-level semantics or active control of generated detail. As an example, FreeU adapts high-to-low frequency ratios in U-Net skip connections dynamically, and operates during inference without additional training. Likewise, at this level, spectral filtering can be employed to disentangle domain-invariant features from style-related spectral noise, thereby facilitating the learning of robust representations from synthetic data. Lastly, architecture-level studies like F-Net and AFNO replace self-attention with frequency-domain operators. The strategy involves the use of global receptive fields to trade off between efficiency in model construction and semantics [12].

Lastly, the Model and Latent Space Level emphasise interventions at the bottleneck or the operator underlying level. The architecture optimization is shifting from peripheral auxiliary modules toward the fundamental reconstruction of core operators. As an example, AFNO does not use spatial-domain self-attention but frequency-domain token mixing. This method also substantially improves computational speed without compromising the global receptive field. It has been extended to a multi-scale frequency-domain mixing mechanism as FreqFormer. It preserves the same linear complexity as single-global filters, but evades the constraints of single global filters on the modeling of local detail [13]. Moreover, the WDNO integrates spectral decomposition directly into the generative bottleneck, producing high-fidelity training data more efficiently. As far as the latent space interventions are concerned, FreeU conducts research on dynamic latent representation $z$ modulation during diffusion. It improves the quality of generation and does not retrain the model by boosting high and low frequencies in the latent space. Lastly, a new paradigm is brought by model-level frequency-domain interventions to provide Privacy-constrained Source-Free Domain Adaptation (SFDA) and Unsupervised Model Adaptation (UMA). Statistical anchors such as frequency-domain statistics are used in studies such as RDA to drive strong parameter calibration in the absence of source-domain data. Such a transition of fixed alignment to dynamical intervention is helpful in building effective and safe cross-domain generative mechanisms.

### 2.3 Cross-Hierarchical Evolution Amidst Frontier Trends

Recent progress with diffusion-based models and large-scale vision models has presented different cross-hierarchical parts of feature coupling. Namely, one algorithm is frequently used in representation spaces to achieve global optimization.

This cross-hierarchical characteristic in diffusion models takes the form of a transition between inference- time plugins and training-side constraints. FreeU shows that the sampling trajectory in latent space can be steered by modifying the skip connections in the feature space.

The frequency-domain intervention is taking a paradigm shift within the landscape of large model research. This shift is one that no longer entails replacing individual operators and rather involves reconstructions of multi-scale architectures. In AFNO, frequency-domain operators were used to recreate token mixing at a fundamental level. FreqFormer later came up with an architectural breakthrough through multi-scale frequency-domain modeling. This is a tendency in which the change from local feature patching to the foundational operator innovation confirms the frequency-domain view as one of the key directions to address the computational bottlenecks. It is also an indication that frequency-domain techniques have been well embedded in the underlying architecture of deep learning systems.

## 3 Performance Evaluation and Analysis of Mainstream Algorithms

**Table 1.** Comparative Analysis of Mainstream Frequency-Domain Image Generation and Adaptation Algorithms.

| Representative Algorithms | Operational Mechanisms | Transfer Levels | Core Transformation Techniques | Key Challenges Addressed | Limitations |
|---|---|---|---|---|---|
| Focal Frequency Loss | Frequency-domain Constraint | Input Space | DFT | Mitigating common "checkerboard" artifacts in generated images | Limited suppression of non-periodic noise |
| FSDR | Frequency Randomization | Feature Space | DFT | Enhancing generalization against non-semantic fluctuations (e.g., weather conditions) | Risk of damaging fine-grained semantics via random perturbations |
| WDNO | Spectral Decomposition | Model and Latent Space | DWT | Structural fidelity and resolution of synthetic datasets for robust AI training | Complex synchronization across different wavelet sub-bands during the generative process |
| GFNet | Frequency-domain Filtering | Model and Latent Space | Learnable Global Filters | Achieving global receptive field with $O(N \log N)$ complexity | Linear growth of frequency-domain parameters with resolution |
| RDA | Spectral Mixing / Adversarial Perturbation | Model and Latent Space | DFT | Enhancing model robustness in Source-Free Domain Adaptation (SFDA) | Relatively high training costs for adversarial defense |
| AFNO | Frequency-domain Filtering / Attention | Model and Latent Space | Fourier Neural Operator (FNO) | Computational overhead of large-scale models processing | High requirements for low-level hardware optimization |

| | | | | | |
|---|---|---|---|---|---|
| | | | | high-resolution images | in operator reconstruction |
| FreeU | Frequency-domain Filtering | Feature/Latent Space | DFT | Improving diffusion model generation quality without retraining | Manual calibration of frequency adjustment ratios for specific scenarios |
| FreqFormer | Frequency-domain Filtering / Attention | Model and Latent Space | Multi-scale Frequency-domain Operators | Insufficient modeling of local textures by global operators | Elevated architectural design complexity |
| Med-D3CG | Spectral Decomposition | Model and Latent Space | DWT and Difference Domain Modeling | Cross-modality medical data synthesis for AI training; improving diffusion efficiency | Requires high-quality paired multi-modality datasets for difference modeling |

A comparative study of these algorithms supports the logic distinctiveness of the frequency-domain paradigm of cross-domain generation (Table 1). This paradigm is further explored in the discussion that follows in terms of three dimensions, namely, evolutionary trajectory, architectural mechanisms, and application paradigms.

To begin with, frequency-domain approaches have diversified out of the external data layer over to the internal semantic layer. This move shows a deeper integration across the levels of transfer and alignment strategies. Early methods focused on global amplitude spectrum exchange in the input space. The practices such as RDA, were later established to establish the significance of feature-space intervention and frequency-band selective enhancement. Recent advancements, such as spectral decoupling frameworks, build upon these principles to achieve granular disentanglement of style and structure. By fully decoupling amplitude and phase components, these models maintain rigorous semantic consistency even amidst extreme stylistic shifts, which is pivotal for generating high-diversity synthetic training sets.

In the architecture mechanisms, the paradigm of frequency-domain has realized a generation leap. This shift towards deterministic transformations takes place towards reconstructing a multi-scale operator. The multi-scale strategy, now advanced by WDNO via wavelet transforms, gives the model a robust foundation in physical interpretability. On the efficiency of the operators, GFNet confirmed the effectiveness of learnable global filtering. At the same time, AFNO with Fourier Neural Operators decreased computational complexity to $O(N \log N)$ . This innovation eliminates cases of redundancy as large-scale architectures are used to handle high-resolution images. Trying to eliminate the shortcomings of single global filters in the representation of local information, FreqFormer uses multi-scale frequency-domain operators. This represents the formal maturation of frequency-domain approaches as anything other than plug-in auxiliaries into more substantial engines in support of massive vision models.

Moreover, the interpretability of frequency-domain features creates a closed loop between synthesis fidelity and domain reliability. Early research, such as Focal Frequency Loss, focused on the direct mitigation of generative artifacts by narrowing the spectral gap between synthetic and real-world distributions. This ensured that generated samples possessed the requisite high-frequency details for robust downstream training. New diffusion-based approaches, exemplified by WDNO, take this intervention to the latent space. It extends intervention to the latent space by proposing a spectrum-aware loss function. This shift is away as an effort at artifact elimination to spectral energy distribution restoration. Together with the flexibility of FreeU and the privacy-sensitive nature of FSDR, these technologies increase the generative robustness and the value of generative deployment to industrial applications. Particularly in precision-heavy fields, Med-D3CG extends these paradigms to medical data synthesis, ensuring cross-modality reliability through wavelet-based difference domain modeling [14].

## 4 Current Limitations

Even though frequency-domain approaches are seen to excel in terms of efficiency, and feature decoupling. These have inherent technical bottlenecks that limit their use.

To begin with, it is difficult to achieve geometric equivariance and impose physical constraints in phase data. Representative methods such as the RDA, are mainly aimed at matching amplitude spectra. This method does not capture the topological and geometrical structural data contained in phase spectra in a sensitive way. Even advanced spectral decoupling frameworks remain vulnerable to semantic distortions triggered by minor frequency-domain interventions. Consequently, such models often struggle to maintain pixel-level spatial consistency when subjected to extreme non-rigid geometric transformations, as global spectral operations lack the inherent geometric equivariance found in spatial convolutions.

Second, semantic frequency-band partitioning is associated with a dilemma of heuristics and ringing effect. Algorithms such as FSDR and FreeU are greatly dependent on frequency thresholds that have been designed manually. Despite the fact that FreqFormer is processing multi-scale frequency-domain operators, the choice of the filter weights is still premised on empirical considerations. Frequency-band intervention used improperly tends to create so-called ringing artifacts at image edges. The phenomenon becomes a critical threat to medical image analysis when the precision requirements are extremely high. At the same time, frequency-domain operators are global, which compromises the possibility of the model to detect fine-grained local details. It is still a problem to trade off the "global semantic consistency" against the "sensitivity to local details" to achieve consistency in the context of complex noise distributions. This leads to a precarious balance of global superiority and local inferiority with the frequency-domain paradigm.

Moreover, the frequency-domain methods have weak generalization on non-natural distributed data. Examples are medical ultrasound and remote sensing Synthetic Aperture Radar (SAR). These data types deviate from the $1/f$ power-law energy distribution which is typical of natural images. Therefore, loss can be considered to have a significant performance degradation similar to Focal Frequency Loss. The tools are mainly made on the basis of standard spectral priors that are not applicable in these areas. Recent specialized frameworks, such as Med-D3CG, have begun to address these challenges by modeling in the wavelet-based difference domain, yet a unified prior for non-natural distributions remains elusive.

At last, there remain difficulties with ensuring the compatibility of architectural refactoring with the underlying hardware. The computational complexity of GFNet, AFNO, FreqFormer and WDNO is theoretically $O(N \log N)$ , which represents a significant efficiency gain over traditional spatial-domain mechanisms. However, practical throughput often deviates from these theoretical bounds depending on the specific hardware implementation and kernel optimization. Hardware is important in ensuring performance based on optimization of the Fast Fourier Transform (FFT). Optimization bottlenecks in Tensor Cores in industrial implementations or on edge devices are still evident. Such limitations require specialized operator libraries to be created.

## 5 Future Outlook

The current direction of frequency-domain studies is towards an even greater paradigm shift as foundation models keep advancing.

First, the state of research will shift to transcending simplistic amplitude alignment in favor of profound phase modelling. The future research needs to fix the problems with semantic distortion in current spectral decoupling frameworks, even in the extreme cases of geometric deformations. Some of the possible avenues are adding phase consistency constraints or building native complex-valued neural networks. Such developments will allow a finer granularity of pixel-level structural integrity. Second, the dynamic frequency management, which is task-driven, will be used instead of heuristic partitioning. Frequency-based attention mechanisms can also be used to further optimize multi-scale operators such as the ones used in FreqFormer and WDNO. By creating systems that can automatically detect and maximize frequency features of critical importance, the ringing effects will be reduced and overall generalization resilience will be increased.

Also, frequency-domain methods have provided an essential route towards Vision-Language Models (VLMs) optimization. By applying architectural-level multi-scale reconstruction using AFNO or FreqFormer, with the addition of a set of inference-time, training-free modulators, it is possible to enhance the quality of generation. This method justifies the use of the industrial application of the "Sustainable AI" paradigm.

Lastly, next-generation structures will have an extensive level of quality consideration and security assurance. Researchers can mitigate generative artifacts by applying spectral energy restoration in the latent space and introducing calibration as part of a generation-in-the-loop paradigm. This combination will also give an opportunity to develop reliable and traceable generative AI systems.

## 6 Conclusion

In this paper, a systematic review of the literature on image generation and cross-domain transfer is described from a frequency-domain perspective. This paper has developed a constructive classification system through an extensive discussion on the Fourier and wavelet transforms. This framework includes five fundamental functional mechanisms and three levels of transfer including input space, feature space and model / latent space.

The frequency-domain methods take advantage of the natural decoupling and global modeling to solve spatial-domain bottlenecks. These techniques excel in modeling long-range dependencies, noise cancellation and balancing cross-domain distributions. The frequency-domain paradigm is distinctly evolutionary. It evolves beyond the external data alignment to feature-level disentanglement and calibration. Such a path will lead to the restoration of latent space energy and multi-scale architecture reconstruction. Also, frequency-domain features enhance generative fidelity and efficient inference. Such benefits facilitate the creation of robust and sustainable AI.

To conclude, the frequency-domain perspective is a paradigm of efficient operation as compared to the traditional spatial-domain counterparts. Difficulties now exist in explicit modeling and precise delineation of semantic boundaries. Nevertheless, architecture-based reconstruction and adaptive modulation have tremendous potential to be explored. In fact, frequency-domain techniques will continue to become one of the driving forces. They will form the basis of the future generation of general-purpose visual representation.

## References


[1] Isola, P., Zhu, J. Y., Zhou, T., Efros, A. A.: Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1125-1134 (2017)

[2] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. et al.: Generative adversarial nets. Advances in neural information processing systems. Vol. 27, pp. 2672-2680 (2014)

[3] Yang, Y., Soatto, S.: Fda: Fourier domain adaptation for semantic segmentation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4085-4095 (2020)

[4] Huang, J., Guan, D., Xiao, A., Lu, S.: Rda: Robust domain adaptation via fourier adversarial attacking. In Proceedings of the IEEE/CVF international conference on computer vision. pp. 8988-8999 (2021)

[5] Yu, Y., Zhan, F., Lu, S., Pan, J., Ma, F., Xie, X., Miao, C.: Wavefill: A wavelet-based generation network for image inpainting. In Proceedings of the IEEE/CVF international conference on computer vision. pp. 14114-14123 (2021)

[6] Huang, J., Guan, D., Xiao, A., Lu, S.: Fsdr: Frequency space domain randomization for domain generalization. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 6891-6902 (2021)

[7] Jiang, L., Dai, B., Wu, W., Loy, C. C.: Focal frequency loss for image reconstruction and synthesis. In Proceedings of the IEEE/CVF international conference on computer vision. pp. 13919-13929 (2021)

[8] Hu, P., Wang, R., Zheng, X., Zhang, T., Feng, H., Feng, R., ... & Wu, T.: Wavelet Diffusion Neural Operator. In International conference on learning representations. (2025)

[9] Guibas, J., Mardani, M., Li, Z., Tao, A., Anandkumar, A., Catanzaro, B.: Adaptive Fourier Neural Operators: Efficient Token Mixers for Transformers. In International conference on learning representations. (2022)

[10] Rao, Y., Zhao, W., Zhu, Z., Lu, J., Zhou, J.: Global filter networks for image classification. Advances in neural information processing systems. 34, 980-993 (2021)
[11] Si, C., Huang, Z., Jiang, Y., Liu, Z.: Freeu: Free lunch in diffusion u-net. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4733-4743 (2024)
[12] Lee-Thorp, J., Ainslie, J., Eckstein, I., Ontanon, S.: Fnet: Mixing tokens with fourier transforms. In Proceedings of the 2022 Conference of the north American chapter of the Association for Computational Linguistics: human language technologies. pp. 4296-4313 (2022)
[13] Dai, T., Wang, J., Guo, H., Li, J., Wang, J., Zhu, Z. : FreqFormer: Frequency-aware Transformer for Lightweight Image Super-resolution. In IJCAI. pp. 731-739 (2024)
[14] Zhu, G., Wan, M., Cao, W., Yu, Z., Hu, J., Li, B., & Fan, X.: Med-D3CG: Wavelet-Based Diffusion in the Difference Domain for Cross-Modality Medical Image Generation. Expert Systems with Applications, 129651 (2025)