ITM Web of Conferences **80**, 01006 (2025) https://doi.org/10.1051/itmconf/20258001006
ACAAI 2025

Optimizing CycleGAN in the Frequency Domain for Artifact Reduction and Detail Preservation

Chang Liu[1*]

[1]School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, Shenzhen 518172, China

Abstract. Unsupervised image transformation models such as CycleGAN have been widely applied. However, they often generate spectral distortions that cause high-frequency artifacts and unrealistic textures. To attempt to address this critical defect, this paper proposes a frequency-aware optimization framework, introducing the spectral consistency loss into the generation process. The key point is the spectral consistency loss, which can minimize the loss of frequency information details and prevent the generation of artifacts. According to the results of experiments, the frequency-aware model reduces the FID of the generated images on the "horse2zebra" benchmark dataset by 15.3% and the KID by 30.1%. In the qualitative comparison, it can be observed that the unnatural artifacts have been significantly suppressed, and spectral analysis also indicates that its frequency distribution is closer to the real image than that of the baseline model. In addition, the human perception survey questionnaire shows that the frequency perception model has won the preference of the majority of people. This work demonstrates that the frequency-domain regularization method is a reliable and effective strategy, which can significantly improve the fidelity and reduce the loss of frequency information for unpaired image transformation, providing a promising approach to mitigating spectral inconsistency in generative models.

1 Introduction

In recent years, Generative Adversarial Networks (GANs) have developed rapidly from the simple content generators to widely used cross-domain transformers in the computer vision field [1]. As a typical representative in unpaired image-to-image transformation area, CycleGAN [2] has made particularly significant contributions undoubtedly. Their applications in science and practice, such as medical imaging [3], have received widespread attention. There are also some studies dedicated to optimizing the quality and authenticity of the generated results by these models [4].

However, the systematic spectral bias in the frequency domain of the methods based on GANs remains a hard challenge that cannot be ignored. The generated images, although spatially conforming to the structure of the source domain, often exhibit serious spectral distortions compared to the real images [5]. This distortion usually manifests itself in the

* Corresponding author: 122090328@link.cuhk.edu.cn

form of missing details and unnatural artifacts, which are the key factors for visual realism. The current standard models are unable to perfectly simulate the spectral characteristics of real and natural images, and this is precisely the limitation of current technologies.

To try and fill this critical gap, this paper proposes a frequency-aware framework to optimize the training of unpaired image-to-image translation models, particularly CycleGAN. The primary contributions of this work are:

- This work proposes a novel spectral loss function that acts as an explicit regularizer for the generator, directly addressing a key limitation of standard models by mitigating spectral distortion.
- It introduces a hybrid reference mechanism to constrain the spectral profile of a generated image, effectively balancing the preservation of structural content from the source with the acquisition of textural style from the target domain.
- The effectiveness of the proposed method is validated through extensive experiments, which demonstrate marked qualitative improvements in visual fidelity and significant quantitative superiority over baseline models.

The remainder of this paper is organized as follows: Section 2 reviews related work. Section 3 details the proposed methodology. Section 4 presents the experimental results, and Section 5 concludes the paper.

2 Related Work

Unpaired Image-to-Image Translation. The field of image-to-image translation evolved from supervised methods like Pix2Pix, which require paired training examples [6], to the unsupervised paradigm established by CycleGAN [2]. Following this breakthrough, a substantial body of work has focused on architectural refinements. For instance, mask-guided mechanisms have been proposed to direct the generator's focus towards salient image regions, thereby enhancing structural preservation [7]. The need for such improvements is underscored by the adoption of these models in high-stakes applications, such as data augmentation for COVID-19 diagnosis systems, where the quality and explainability of generated images are paramount [8]. While these spatial-domain enhancements are valuable, they do not directly address the underlying problem of spectral information loss.

Architectural Sources of Generation Artifacts. It is now well-established that the architectural components of deep neural networks can themselves be a source of detectable artifacts. Transposed convolutions, a common up-sampling method in generator networks, are widely known to cause characteristic checkerboard patterns in the output image [9]. More generally, the entire convolutional process creates unique, traceable patterns that can betray an image's synthetic origin [10]. These architectural limitations highlight that artifacts are not merely a product of training instability but are woven into the fabric of the generators themselves, leaving distinct signatures in the high-frequency spectrum of the output image.

Frequency-Domain Analysis of Synthetic Images. The aforementioned artifacts become particularly evident when images are analyzed in the frequency domain. As surveyed by Tolosana et al. [11], the detection of sophisticated manipulations like deepfakes frequently relies on exposing inconsistencies in the frequency domain, confirming that these spectral artifacts represent a systemic and exploitable vulnerability. Another powerful frequency-based approach involves analyzing high-frequency noise. Generative models often fail to replicate the unique, camera-specific noise patterns or "noiseprints" present in real images, providing a robust method for detecting synthetic content [12]. The success of these detection methods underscores a key vulnerability in current generative models and indicates that a direct intervention in the frequency domain is necessary to correct the generator's inherent spectral distortion.

3 Methodology

This section will provide a detailed introduction to the frequency-aware optimization method based on the CycleGAN architecture, including the methodology of spectral features, mixed reference benchmarks, and spectral consistency loss.

3.1 Cycle-Consistent Adversarial Networks (CycleGAN)

The method proposed in this paper is based on the CycleGAN framework [2], and the essence of CycleGAN is a kind of GAN models that can achieve unsupervised image transformation. For a standard GAN, it includes a generator G that maps latent vectors $\mathbf{z}$ to generated images, and a discriminator D that distinguishes between generated images and real images. This adversarial process is formulated as the minimax objective:

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)}[\log D(x)] + E_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z})}\left[\log\left(1 - D\big(G(\mathbf{z})\big)\right)\right] \tag{1}$$

CycleGAN extends the principles of GAN, achieving the back-and-forth transformation between the source domain A and the target domain B by using two pairs of generators and detectors. The key point is that it introduces the cycle consistency loss (L_{cyc}), which enforces the constraint that the round-trip transformation can restore the original image. This constraint is expressed as:

$$L_{cyc}(G_{AB}, G_{BA}) = E_{a \sim p_{data}(A)}\left[|G_{BA}\big(G_{AB}(a)\big) - a|_1\right] + E_{b \sim p_{data}(B)}\left[|G_{AB}\big(G_{BA}(b)\big) - b|_1\right] \tag{2}$$

The complete objective function for CycleGAN integrates the adversarial losses with this cycle-consistency loss, weighted by a hyperparameter λ:

$$L(G_{AB}, G_{BA}, D_A, D_B) = L_{GAN}(G_{AB}, D_B) + L_{GAN}(G_{BA}, D_A) + \lambda L_{cyc}(G_{AB}, G_{BA}) \tag{3}$$

While this formulation is highly effective, its exclusive focus on spatial-domain objectives can result in the generation of images with spectral inconsistencies, manifesting as high-frequency artifacts. This limitation motivates the frequency-aware method proposed herein.

3.2 Frequency-Aware Spectral Optimization

To imbue the CycleGAN framework with frequency-domain awareness, a novel spectral optimization strategy is introduced. This strategy is realized through three core components: a method for extracting a spectral fingerprint, a hybrid reference mechanism for defining a target spectrum, and a tailored loss function to enforce spectral consistency.

3.2.1 Spectral Fingerprint

The primary analytical tool is the radially-averaged Power Spectral Density (PSD), which provides a compact, rotation-invariant signature of an image's frequency content. For a two-dimensional image I, its Fourier Transform $F(I)$ is first computed. The power spectrum is then given by $|F(I)|^2$. This 2D spectrum is converted into a 1D profile by averaging power values within concentric annular bins, indexed by their radial distance from the frequency spectrum's center. The resulting 1D PSD vector is log-scaled for stability and then normalized via a Softmax function to produce a probability distribution, denoted as $\mathbf{p}_{\mathbf{psd}}$:

$$\mathbf{p_{psd}}(k) = \frac{\exp(\log(\mathrm{PSD}(k)))}{\sum_{j=1}^{N_{\mathrm{bins}}} \exp(\log(\mathrm{PSD}(j)))} \tag{4}$$

This distribution $\mathbf{p_{psd}}$ serves as a characteristic spectral fingerprint for the image.

3.2.2 Hybrid Reference Baseline

Defining an appropriate target spectrum is a critical challenge. Simply forcing a generated image G_{AB} (a) to precisely match the average spectrum of the target domain $\boldsymbol{\mu}_{\mathbf{ref}}$, could inadvertently suppress essential low-frequency structural information inherited from the source image. To address this, a dynamic and hybrid reference baseline $\mathbf{p_{baseline}}$ is proposed. This baseline is formulated as a weighted interpolation between the spectral profile of the source image, $\mathbf{p_{psd}}(a)$, and the pre-computed mean spectral profile of the target domain, $\boldsymbol{\mu}_{\mathbf{ref}}$:

$$\mathbf{p_{baseline}} = (1 - w) \cdot \mathbf{p_{psd}}(a) + w \cdot \boldsymbol{\mu}_{\mathbf{ref}} \tag{5}$$

The weight w balances the preservation of source content (dominant in low frequencies) with the acquisition of target style (characterized by high frequencies).

3.2.3 Spectral Consistency Loss

With the hybrid baseline established, a loss function is needed to penalize significant deviations. A simple L1 or L2 norm can be overly punitive, penalizing minor spectral fluctuations that may represent desirable variations. Therefore, to afford the generator greater flexibility and to stabilize training, the proposed spectral consistency loss L_{spec} is formulated as an epsilon-insensitive function. This design introduces a tolerance margin around the baseline, defined by the standard deviation of the target domain's spectral profile, $\boldsymbol{\sigma}_{\mathbf{ref}}$, scaled by a sensitivity hyperparameter k. A penalty is incurred only when the absolute difference between the PSD of the generated image, $\mathbf{p_{psd}}$ $(G_{AB}$ $(a))$, and the baseline, $\mathbf{p_{baseline}}$, exceeds this margin. The loss is implemented using the robust Smooth L1 function:

$$L_{spec}(G_{AB}(a)) = E\left[\mathrm{SmoothL1}\left(\max\left(0, \left|\mathbf{p_{psd}}(G_{AB}(a)) - \mathbf{p_{baseline}}\right| - k \cdot \boldsymbol{\sigma}_{\mathbf{ref}}\right)\right)\right] \tag{6}$$

This spectral loss is then integrated into the final generator objective, weighted by a hyperparameter λ_{spec}, which is gradually introduced via a linear warm-up strategy to ensure stable convergence:

$$L_{G_{total}} = L(G_{AB}, G_{BA}, D_A, D_B) + \lambda_{spec} L_{spec}(G_{AB}(a)) \tag{7}$$

4 Experiments

This section will provide a detailed description of the experimental setup and specific details, and will also analyze the results of experiments.

4.1 Experimental Setup

4.1.1 Datasets

A large number of dominant experiments and tests were conducted on the "horse-to-zebra" standard benchmark dataset. To verify the generalization ability, the model was also tested on several other image transformation datasets with different structural and texture features, including “zebra2horse”, “apple2orange”, “orange2apple”, “summer2winter_yosemite”, and “winter2summer_yosemite”.

4.1.2 Implementation Details

The proposed model is built upon the official PyTorch implementation of CycleGAN, which utilizes a ResNet-9 generator and a 70×70 PatchGAN discriminator. The frequency-aware spectral loss, as detailed in Section 3, was integrated into this architecture. For a fair comparison, all models were trained with identical data preprocessing, hyperparameters, and training environments. Key hyperparameters for the proposed spectral loss component are summarized in Table 1. All experiments were performed running Ubuntu 18.04 on NVIDIA V100-32GB GPUs with a batch size of 2. For the proposed frequency-aware models, the spectral loss part was introduced using a linear warm-up strategy, with an effective warm-up duration corresponding to approximately 37.5 epochs.

Table 1. Optimal Hyperparameters for the Proposed Method.

Hyperparameter	Value	Description
λ_{spec}	0.5	Overall weight of the spectral consistency loss.
w	0.3	Weight of the target domain in the hybrid baseline.
k	0.9	Tolerance factor for the epsilon-insensitive spectral loss.

4.1.3 Evaluation Protocol

A multi-faceted evaluation protocol was adopted to provide a holistic assessment of model performance.

Quantitative Metrics: The Fréchet Inception Distance (FID) and Kernel Inception Distance (KID) were used to measure the statistical similarity between the distributions of generated and real images. Lower scores indicate higher quality and diversity.

Spectral Profile Analysis: To directly evaluate the frequency-domain regularization, the 1D radially-averaged Power Spectral Density (PSD) was computed for all test images. This allows for a direct comparison of the spectral characteristics of real images, baseline outputs, and the proposed model's outputs.

Human Perceptual Research: Human perception has always been the gold standard for evaluating image quality. To assess visual quality, a user study was conducted. Participants were presented with pairs of images generated by the baseline and the proposed model and were asked to select the image with greater performance in a single-blind controlled test.

4.2 Results and Analysis

4.2.1 Hyperparameter Analysis

To determine the optimal configuration for the spectral loss function, an extensive analysis of its key hyperparameters was performed. The investigation focused on the overall loss weight λ_{spec} and the tolerance factor k. These parameters were systematically varied, and their impact on FID and KID scores was evaluated on the "horse2zebra" test dataset.

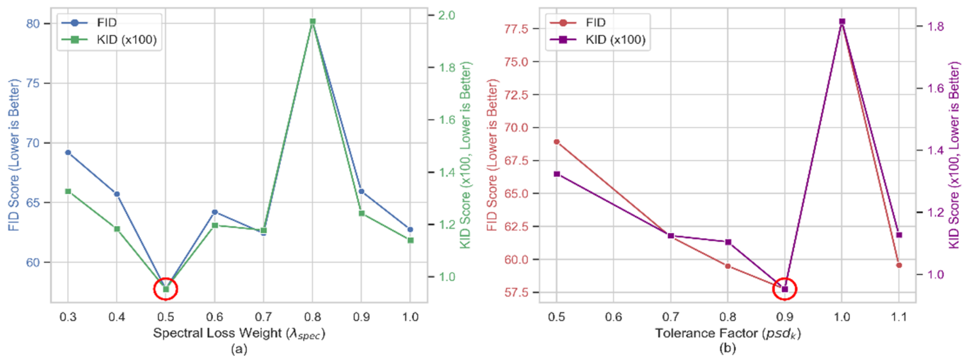

Fig. 1. Ablation study on spectral loss hyperparameters λ_{spec} and k. The impact of each parameter on FID and KID scores on the horse2zebra test set is shown as (a) metrics vs. λ_{spec} and (b) metrics vs. k.

As visualized in Figure 1, the experimental results indicate that both parameters exhibit clear optimal ranges. The analysis for λ_{spec} revealed that a moderate value of 0.5 achieved the best performance. Smaller values failed to provide a strong enough training signal to correct spectral deviations, while larger values over-penalized the generator, causing conflicts with the primary adversarial and cycle-consistency objectives. Similarly, the tolerance factor k showed an optimal value around 0.9. A stricter constraint excessively regularized the generator, stifling plausible variations, while a looser constraint failed to provide sufficient guidance to prevent artifact formation. This analysis led to the selection of the optimal configuration presented in Table 1, which was used for all subsequent experiments.

4.2.2 Main Comparative Results

Quantitative Comparison: As summarized in Table 2, the proposed frequency-aware method consistently and significantly outperforms the baseline CycleGAN on the horse2zebra test dataset. The model achieves a Fréchet Inception Distance (FID) of 57.75, representing a 15.3% reduction compared to the baseline's 68.22. It obtains a Kernel Inception Distance (KID) of 0.95×10^{-2}, a 30.1% improvement over the baseline's 1.36×10^{-2}. These results provide strong quantitative evidence that the integration of the spectral loss leads to a generated image distribution substantially closer to the target domain.

Table 2. Main Quantitative Results on the horse2zebra dataset.

Model	**FID**	**KID (×100)**
CycleGAN (Baseline)	68.22	1.36
Frequency-aware CycleGAN (Ours)	**57.75 (reduced by 15.3%)**	**0.95 (reduced by 30.1%)**

https://doi.org/10.1051/itmconf/20258001006

Qualitative Comparison: The quantitative improvements are strongly corroborated by a qualitative comparison, as shown in Figure 2.

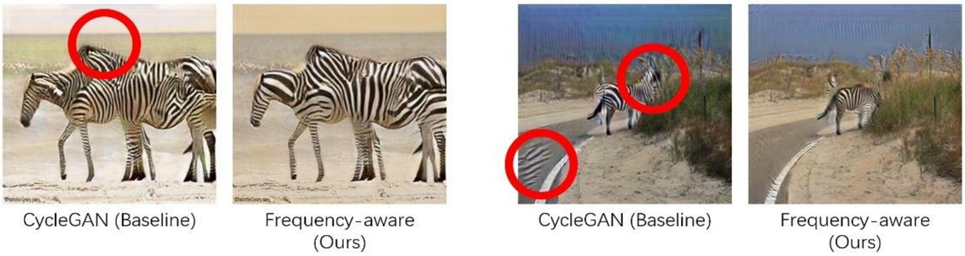

Fig. 2. A parallel qualitative comparison of the "horse2zebra" task. The output of the baseline model shows obvious high-frequency noise or inappropriate zebra patterns on the horse's body, while the output of the frequency-aware model is clearer and more in line with the real situation.

Images produced by the baseline CycleGAN frequently exhibit distracting high-frequency artifacts, such as noisy textures and unnatural ripple patterns, particularly in smooth regions. In stark contrast, the proposed frequency-aware model successfully suppresses these distortions, yielding images that are visibly cleaner and more coherent, and it also demonstrates superior performance in preserving fine structural details from the source domain, avoiding the over-smoothing often observed in the baseline's outputs.

Spectral Profile Analysis: The effectiveness of the proposed spectral loss is directly visualized in the aggregated Power Spectral Density plot.

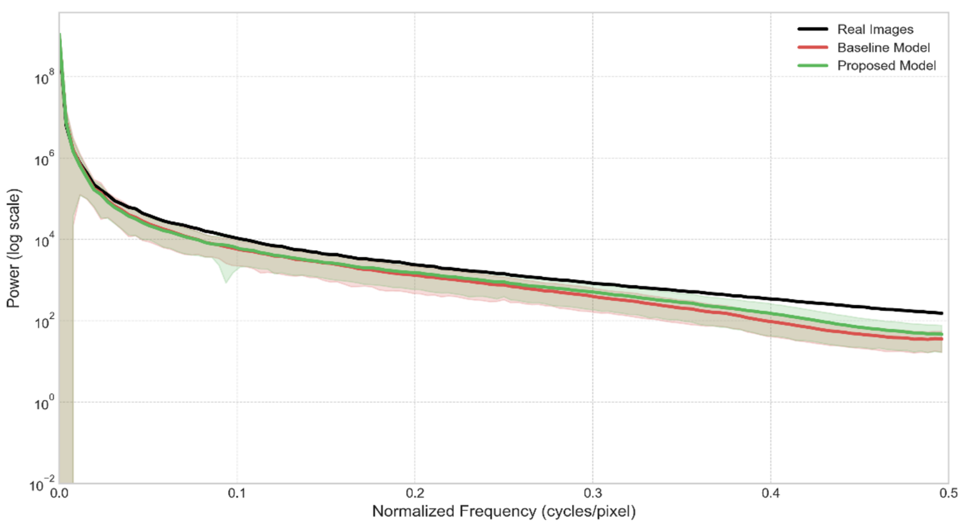

Fig. 3. Average radial average power spectral density (PSD) curve graph. The curve output by the model proposed in this paper is closer to the frequency distribution of the real image compared to the curve of the benchmark model, especially in the mid and high frequency range.

As illustrated in Figure 3, the PSD profile of the baseline model deviates considerably from the ground truth profile of real images, particularly showing a deficit in the mid-to-high frequency range. The wide shaded area also indicates a high variance in its spectral outputs, correlating with the inconsistent appearance of artifacts. In contrast, the curve of the proposed model shows a much closer alignment to the ground truth, successfully capturing the overall spectral trend. This improvement is particularly evident in the mid-to-high frequencies where the baseline fails, demonstrating a marked improvement in generation consistency and spectral stability. This analysis provides strong evidence that the spectral consistency loss successfully regularizes the generator, directly causing the observed improvements in visual fidelity and artifact reduction.

4.2.3 Generalization Validation

To confirm that the method's advantages are not confined to a single dataset, the same trained model configuration was applied to the "zebra2horse", "apple2orange", "orange2apple", "summer2winter_yosemite", and "winter2summer_yosemite" tasks. Figure 4 selects some representative generated images for comparison. Compared with baseline CycleGAN model, the images generated by the frequency-aware model have fewer artifacts and their colors are closer to the style of the target, which is precisely the external manifestation of spectral consistency. The results of quantitative analysis, presented in Figure 5, mirror the findings from the primary experiments. The frequency-aware model again shows significant improvements over the baseline, demonstrating that the proposed frequency-aware optimization is a robust and generalizable enhancement for unpaired image translation.

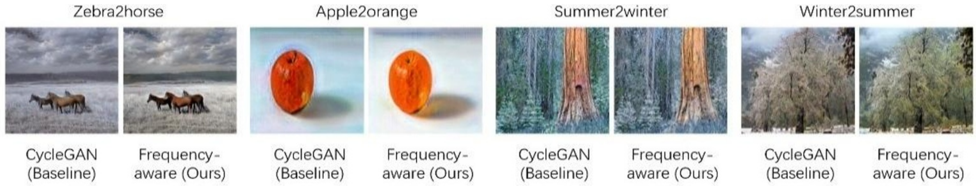

Fig. 4. The qualitative examples in other scenario tasks demonstrate that the frequency-aware model can achieve image denoising processing and better preserve the detailed information.

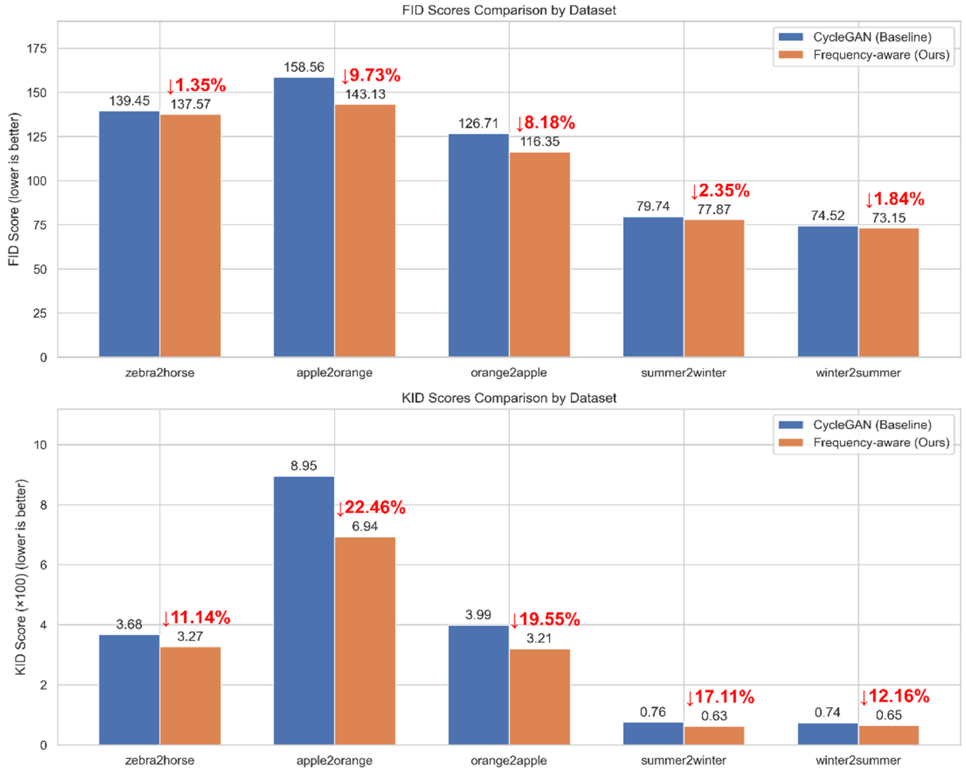

Fig. 5. The frequency-aware model significantly reduced the FID and KID values on multiple datasets, demonstrating the generalization ability of this method.

4.2.4 Human Perceptual Research

The results from the single-blind controlled questionnaire provide definitive evidence of the perceptual superiority of the proposed method.

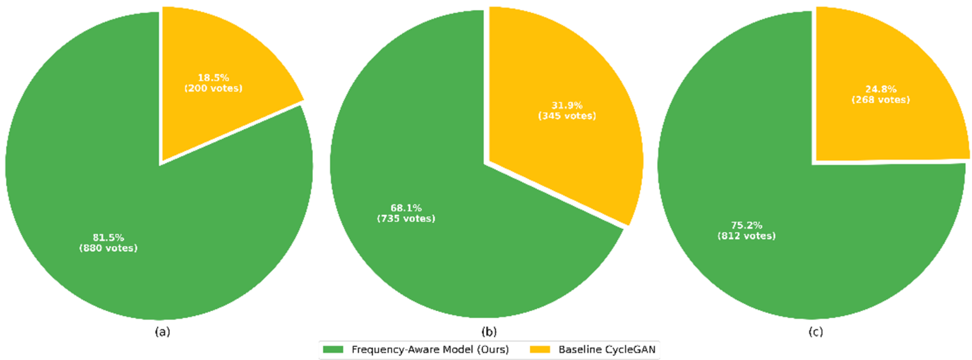

Fig. 6. A pie chart summarizing the user study results shows that users strongly prefer images generated by the frequency-aware model based on (a) fewer visual artifacts, (b) better detail preservation, and (c) overall visual quality.

A total of 54 participants evaluated 20 pairs of translated images each. The results of the questionnaire survey are shown in Figure 6. When asked to identify the image with fewer visual artifacts, participants chose the output from the frequency-aware model in 81.5% of the comparisons (880 out of 1080 total trials). This model was also favored for its ability to preserve fine details, earning 68.1% of the votes (735 out of 1080). For overall visual quality and realism, it was also preferred in 75.2% of the cases (812 out of 1080). These findings confirm that the quantitative improvements achieved by this model translate into a meaningfully better and more reliable visual experience for human observers.

5 Conclusion

This paper introduces a frequency-aware optimization framework for unpaired image-to-image translation, addressing the spectral inconsistencies inherent to the conventional CycleGAN. By integrating a spectral consistency loss, the proposed method achieves significant improvements in generation quality, effectively suppressing the high-frequency artifacts common in standard adversarial training.

The framework's efficacy was validated through comprehensive experiments. On the primary benchmark task, the model achieved a 15.3% reduction in FID and a 30.1% reduction in KID over the baseline. These quantitative gains were strongly corroborated by human perceptual research, where participants significantly preferred the proposed model's outputs for artifact reduction (81.5% preference), detail preservation (68.1%), and overall realism (75.2%). While the proposed method introduces a moderate computational overhead during training, the resulting improvements in image quality justify this cost, and the inference-time efficiency remains unaffected.

Future work could extend this framework to higher-resolution synthesis, adapt it to other generative architectures beyond CycleGAN, and explore more computationally efficient formulations of the spectral loss. Ultimately, the proposed frequency-aware approach provides a promising direction for integrating spectral constraints into adversarial training, opening new avenues for enhancing the fidelity of generative image translation.

References

1. M. Y. Liu, X. Huang, J. Yu, T. C. Wang, A. Mallya, Generative adversarial networks for image and video synthesis: Algorithms and applications. Proc. IEEE **109**, 839-862 (2021) https://doi.org/10.1109/JPROC.2021.3049196
2. J. Y. Zhu, T. Park, P. Isola, A. A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, Oct 22-29 (2017), pp. 2223-2232, https://doi.org/10.1109/ICCV.2017.244
3. S. Kazeminia, C. Baur, A. Kuijper, B. Van Ginneken, N. Navab, S. Albarqouni, A. Mukhopadhyay, GANs for medical image analysis. Artif. Intell. Med. **109**, 101938 (2020) https://doi.org/10.1016/j.artmed.2020.101938
4. L. Chen, L. Wu, Z. Hu, M. Wang, Quality-Aware Unpaired Image-to-Image Translation. IEEE Trans. Multimedia **21**, 2664-2674 (2019) https://doi.org/10.1109/TMM.2019.2907052
5. R. Durall, M. Keuper, J. Keuper, Watch your up-convolution: Cnn based generative deep neural networks are failing to reproduce spectral distributions. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 7890-7899 (2020) https://doi.org/10.48550/arXiv.2003.01826
6. P. Isola, J. Y. Zhu, T. Zhou, A. A. Efros, Image-to-image translation with conditional adversarial networks, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2017), pp. 1125-1134, https://doi.org/10.1109/CVPR.2017.632
7. G. Hu, Y. Zheng, H. Yan, G. Hua, Y. Yan, Mask-guided cycle-GAN for specular highlight removal. Pattern Recognit. Lett. **161**, 108-114 (2022) https://doi.org/10.1016/j.patrec.2022.06.014
8. Y. H. Wu, S. H. Gao, J. Mei, J. Xu, D. P. Fan, R. G. Zhang, M. M. Cheng, Jcs: An explainable covid-19 diagnosis system by joint classification and segmentation. IEEE Trans. Image Process. **30**, 3113-3126 (2021) https://doi.org/10.1109/TIP.2021.3058783
9. A. Odena, V. Dumoulin, C. Olah, Deconvolution and checkerboard artifacts. Distill **1**, e3 (2016) http://doi.org/10.23915/distill.00003
10. L. Guarnera, O. Giudice, S. Battiato, Fighting deepfake by exposing the convolutional traces on images. IEEE Access **8**, 165085-165098 (2020) https://doi.org/10.1109/ACCESS.2020.3023037
11. R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, J. Ortega-Garcia, Deepfakes and beyond: A survey of face manipulation and fake detection. Inf. Fusion **64**, 131-148 (2020) https://doi.org/10.1016/j.inffus.2020.06.014
12. D. Cozzolino, L. Verdoliva, Noiseprint: A CNN-based camera model fingerprint. IEEE Trans. Inf. Forensics Secur. **15**, 144-159 (2019) https://doi.org/10.1109/TIFS.2019.2916364