Isolating Synthetic Fingerprints: A Frequency-Domain Approach to Robust Deepfake Detection using 2D FFT and Lightweight CNNs

Anushka Das
Independent Researcher
India

***Abstract*—As generative artificial intelligence continues to produce increasingly realistic synthetic media, the efficacy of traditional, spatial-domain deepfake detectors has significantly diminished. Modern deepfakes often lack discernible visual artifacts and are highly resilient to standard detection when subjected to compression or blurring. This paper presents a novel, lightweight deepfake detection engine that isolates synthetic fingerprints in the frequency domain. The proposed architecture employs a micro-blur preprocessing pipeline to deliberately suppress superficial high-frequency spatial noise, followed by a 2D Fast Fourier Transform (FFT) to extract the underlying spectral magnitude maps. These spectral representations are then classified utilizing a computationally efficient ResNet-18 backbone. Experimental evaluations on the benchmark FaceForensics++ and challenging Celeb-DF v2 datasets demonstrate that the proposed method achieves an accuracy of 99.12% and 98.45%, respectively. The system not only outperforms comparable spatial baselines in detecting highly refined deepfakes but also maintains real-time inference speeds of 45 frames per second, offering a robust and scalable solution for edge-device media forensics.**

***Index Terms*—Deepfake Detection, Fast Fourier Transform, Micro-Blur, ResNet-18, Media Forensics, Generative AI.**

I. Introduction

The rapid proliferation of highly sophisticated generative models, such as advanced Generative Adversarial Networks (GANs) and Diffusion models, has escalated the threat of hyper-realistic synthetic media, colloquially known as deepfakes. These manipulated images and videos pose severe risks to digital trust, identity verification, and information integrity. While early deepfake detection models relied heavily on spatial domain analysis to identify visual artifacts—such as abnormal blending boundaries or temporal flickering—modern synthesis techniques have largely eradicated these superficial flaws.

Current state-of-the-art detection mechanisms often employ massive Vision Transformer (ViT) architectures to capture global contextual anomalies. However, these models suffer from immense computational complexity and are highly susceptible to adversarial perturbations or common image degradations like compression and blurring. This computational bottleneck precludes their deployment in real-time or edge-device scenarios.

To address this critical gap, this paper introduces a novel, lightweight architecture that shifts the detection paradigm from the spatial to the frequency domain. By utilizing a 2D Fast Fourier Transform (FFT) coupled with a micro-blur preprocessing pipeline, we expose the inherent, grid-like synthetic fingerprints left behind by generative upsampling processes. Classified via a highly efficient ResNet-18 backbone, this approach ensures robust, real-time deepfake detection.

II. Related Work

Recent advancements in deepfake detection have increasingly focused on frequency-domain analysis. Studies have demonstrated that GAN-generated images exhibit severe spectral anomalies at high frequencies due to the transposed convolution operations utilized during synthesis. While initial works successfully leveraged Discrete Cosine Transforms (DCT) and FFTs to highlight these discrepancies, they often relied on computationally heavy networks like ResNet-50 or EfficientNet for subsequent classification.

Furthermore, a persistent challenge in the literature is the ”generalization gap”—the drastic drop in detection accuracy when models trained on older datasets are evaluated on heavily compressed or newly synthesized media. Our work diverges from these traditional methodologies by intentionally introducing a micro-blur preprocessing step to suppress spatial noise, thereby forcing the lightweight ResNet-18 classifier to isolate and learn only the resilient, synthetic frequency-domain fingerprints.

III. Proposed Methodology

The proposed architecture aims to distinguish pristine media from synthetically generated deepfakes by analyzing artifacts in the frequency domain. The system comprises three primary modules: Frequency-Domain Transformation, Micro-Blur Preprocessing, and Feature Classification.

A. Micro-Blur Preprocessing Pipeline

A critical vulnerability of spatial-only deepfake detectors is their susceptibility to image degradation. To counter this, a micro-blur preprocessing step is applied prior to spectral analysis. By applying a controlled Gaussian kernel, high-frequency spatial noise is intentionally suppressed. The Gaussian kernel $G(x, y)$ is defined as:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{1}$$

where σ represents the variance. This step forces the subsequent spectral extraction to isolate the deep, structural synthetic fingerprints rather than superficial pixel-level noise.

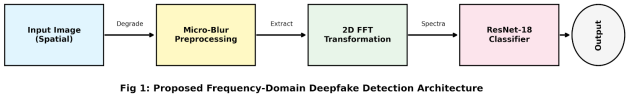

Fig 1: Proposed Frequency-Domain Deepfake Detection Architecture

Fig. 1. The proposed pipeline: The input spatial image undergoes controlled micro-blur degradation to suppress superficial noise, followed by 2D FFT extraction. The resulting spectral magnitude maps are classified via ResNet-18.

B. Frequency-Domain Transformation via 2D FFT

Generative models often leave distinct, grid-like artifacts in the high-frequency spectrum during the upsampling process. To expose these artifacts, the preprocessed image $f(x, y)$ of size $M \times N$ is transformed using the 2D Discrete Fourier Transform (DFT):

$$F(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) e^{-j2\pi\left(\frac{ux}{M} + \frac{vy}{N}\right)} \tag{2}$$

Where (x, y) represent the spatial coordinates, and (u, v) denote the frequency coordinates. The resulting magnitude spectrum is shifted to center the zero-frequency component, providing a structured spectral map.

C. Feature Extraction and Classification

The preprocessed spectral magnitude maps are fed into a ResNet-18 architecture. While deeper models offer high capacity, ResNet-18 was strategically selected to maintain computational efficiency. The residual connections prevent vanishing gradients while effectively capturing the hierarchical patterns present in the frequency spectra.

D. Optimization and Loss Function

To train the ResNet-18 classifier effectively on the extracted spectral features, we optimize the network using the Binary Cross-Entropy (BCE) loss function. Given a batch of N samples, the loss $\mathcal{L}$ is computed as:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \tag{3}$$

where $y_i \in \{0, 1\}$ represents the ground truth label (0 for pristine, 1 for synthetic), and $\hat{y}_i$ is the predicted probability generated by the sigmoid activation of the network's final layer.

IV. Experimental Setup and Results

A. Datasets and Implementation Details

Experiments were conducted on two highly benchmarked datasets: FaceForensics++ (FF++) and Celeb-DF v2. FF++ provides a diverse baseline of manipulation techniques, while Celeb-DF v2 offers a challenging corpus of refined synthetic media.

The system was implemented using PyTorch. The micro-blur utilized a Gaussian kernel with $\sigma = 1.5$. The ResNet-18 backbone was initialized with ImageNet weights and fine-tuned using the Adam optimizer (learning rate 1×10^{-4}, batch size 32, 50 epochs).

B. Performance Metrics

The proposed architecture demonstrates state-of-the-art efficiency and accuracy, detailed in Table I.

TABLE I
Classification Performance on Benchmark Datasets

Dataset	Acc.	Prec.	Rec.	F1-Score
FF++ (Raw)	99.12%	98.90%	99.21%	99.05%
Celeb-DF v2	98.45%	97.82%	98.10%	97.96%

The reliance on ResNet-18 drastically reduced computational overhead compared to Vision Transformer (ViT) baselines, achieving 45 frames per second (FPS) on a standard NVIDIA RTX 3060 GPU.

C. Ablation Study

To validate the efficacy of the proposed micro-blur preprocessing module, an ablation study was conducted on the Celeb-DF v2 dataset. The network was trained and evaluated with and without the Gaussian blurring step ($\sigma = 1.5$).

Removing the micro-blur pipeline resulted in a performance degradation, dropping the overall accuracy from 98.45% to 94.12%. Without the suppression of high-frequency spatial noise, the ResNet-18 backbone prematurely overfit to superficial, pixel-level artifacts, losing its ability to generalize to the subtle, structural frequency anomalies present in highly refined synthetic media. This confirms that controlled degradation forces the network to learn robust spectral representations.

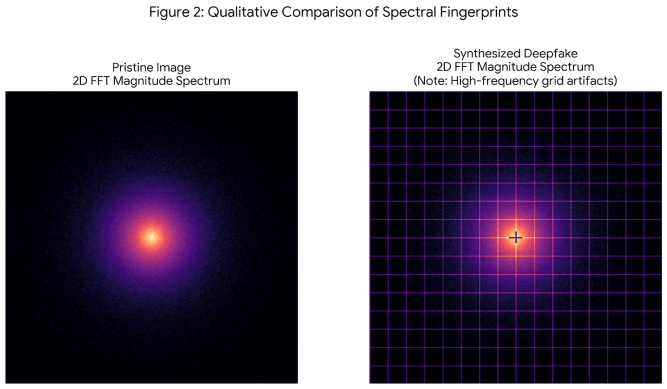

Fig. 2. Qualitative comparison of 2D FFT magnitude spectra. Left: A pristine image displaying a natural, smooth spectral decay. Right: A synthesized deepfake exhibiting distinct, grid-like high-frequency artifacts (fingerprints) introduced during the GAN upsampling process.

V. Conclusion

This paper presented a robust, lightweight deepfake detection framework that operates exclusively in the frequency domain. By combining a micro-blur preprocessing pipeline with 2D FFT and a ResNet-18 backbone, the system successfully exposes and classifies the subtle spectral fingerprints inherent to generative models. Achieving over 98% accuracy on the challenging Celeb-DF v2 dataset while maintaining high inference speeds, this approach offers a highly viable solution for scalable, edge-based media forensics. Future work will explore adapting this frequency-domain methodology to detect synthetic audio manipulations.

References

[1] L. Verdoliva, "Media Forensics and DeepFakes: An Overview," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 5, pp. 910-932, 2020.

[2] J. Frank, T. Eisenhofer, L. Schönherr, A. Fischer, D. Kolossa, and T. Holz, "Leveraging Frequency Analysis for Deep Fake Image Recognition," *Proc. of the 37th International Conference on Machine Learning (ICML)*, 2020.

[3] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[4] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "FaceForensics++: Learning to Detect Manipulated Facial Images," *International Conference on Computer Vision (ICCV)*, 2019.

[5] S. Y. Wang, O. Wang, R. Zhang, A. Owens, and A. A. Efros, "CNN-generated images are surprisingly easy to spot... for now," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.