NEURAL INFORMATION PROCESSING SYSTEMS 2024

FreqMark: Invisible Image Watermarking via Frequency Based Optimization in Latent Space

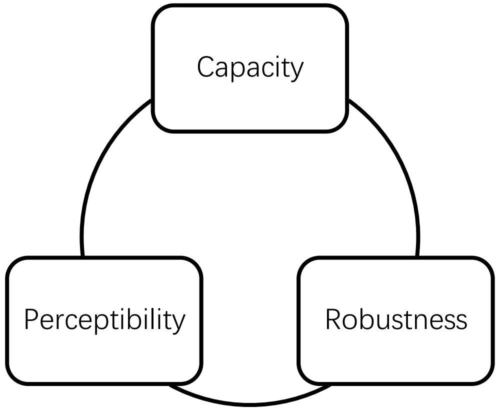
Yiyang Guo[*1, 5 †], Ruizhe Li[*2], Mude Hui[3], Hanzhong Guo[4], Chen Zhang[1]

Chuangjian Cai[5], Le Wan[5], Shangfei Wang[‡1]

[1]University of Science and Technology of China [2]Fudan University
[3]University of California, Santa Cruz [4]The University of Hong Kong [5]IEG, Tencent

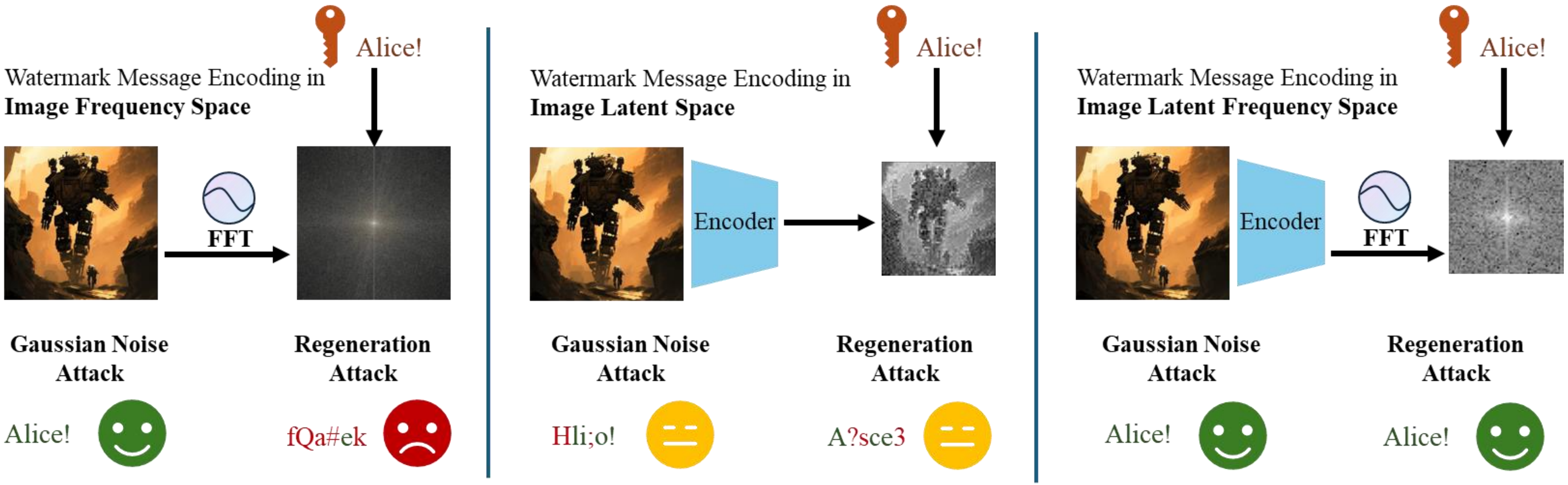
Image Watermarking

- Deepfakes
- Copyright Infringement



Motivation

- Embedding watermarks in the latent frequency space of images.

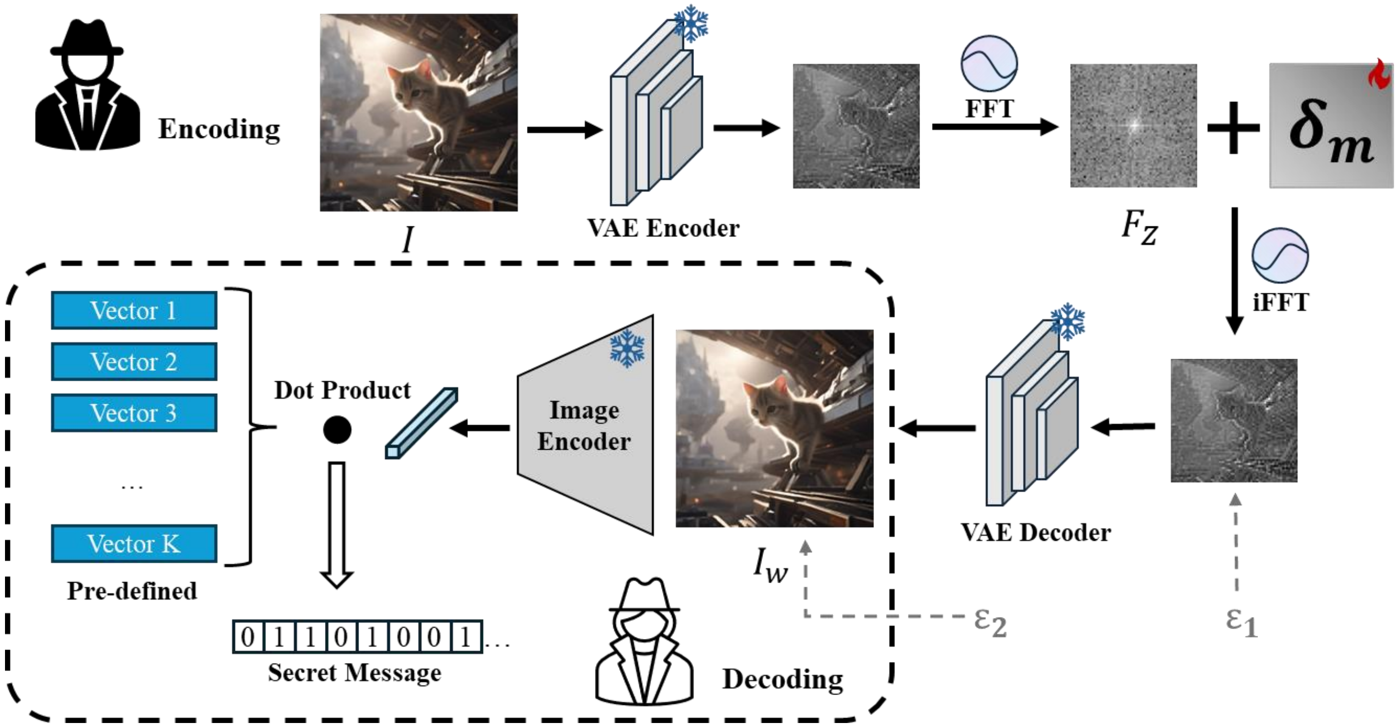


Contributions

- FreqMark encodes hidden messages within the **latent frequency space** of images and achieves watermark embedding through indirect optimization centered on the image itself **without requiring network training**.
- FreqMark is **highly flexible**, allowing for a free trade-off between the bits number of the encoded message, image quality and watermark robustness to meet diverse requirements.
- FreqMark demonstrates **significant robustness advantages**, particularly during regeneration attacks compared to baseline methods.

Flexibility & Robustness

Method



- Encoding

$$I_w = D\left(FFT^{-1}\left(FFT(E(I)) + \delta_m\right)\right)$$

- Decoding

$$m_d^k = sign\left(z_{I_w} \cdot v_k\right) = sign\left(E_{img}(I_w) \cdot v_k\right), v_k \in V_K^N$$

$$V_K^N = \{v_1, v_2, \ldots, v_K \mid K \leq N\}$$ Pre-defined Vectors

Training

- Image Quality

$$\mathcal{L}_p = -PSNR(I_w, I)$$

$$\mathcal{L}_i = LPIPS(I_w, I)$$

- Watermark Message

$$\mathcal{L}_m(I_w) = \frac{1}{K}\sum_{k=1}^{K} max(0, (\mu - (z_{I_w} \cdot v_k) \cdot m_k)) \,, v_k \in V_K^N, m_k \in \{-1,1\}$$

- Robustness Enhancement

$$I_{p1} = D(FFT^{-1}(F_Z + \delta_m) + \epsilon 1)$$

$$I_{p2} = D\big(FFT^{-1}(F_Z + \delta_m)\big) + \epsilon 2$$

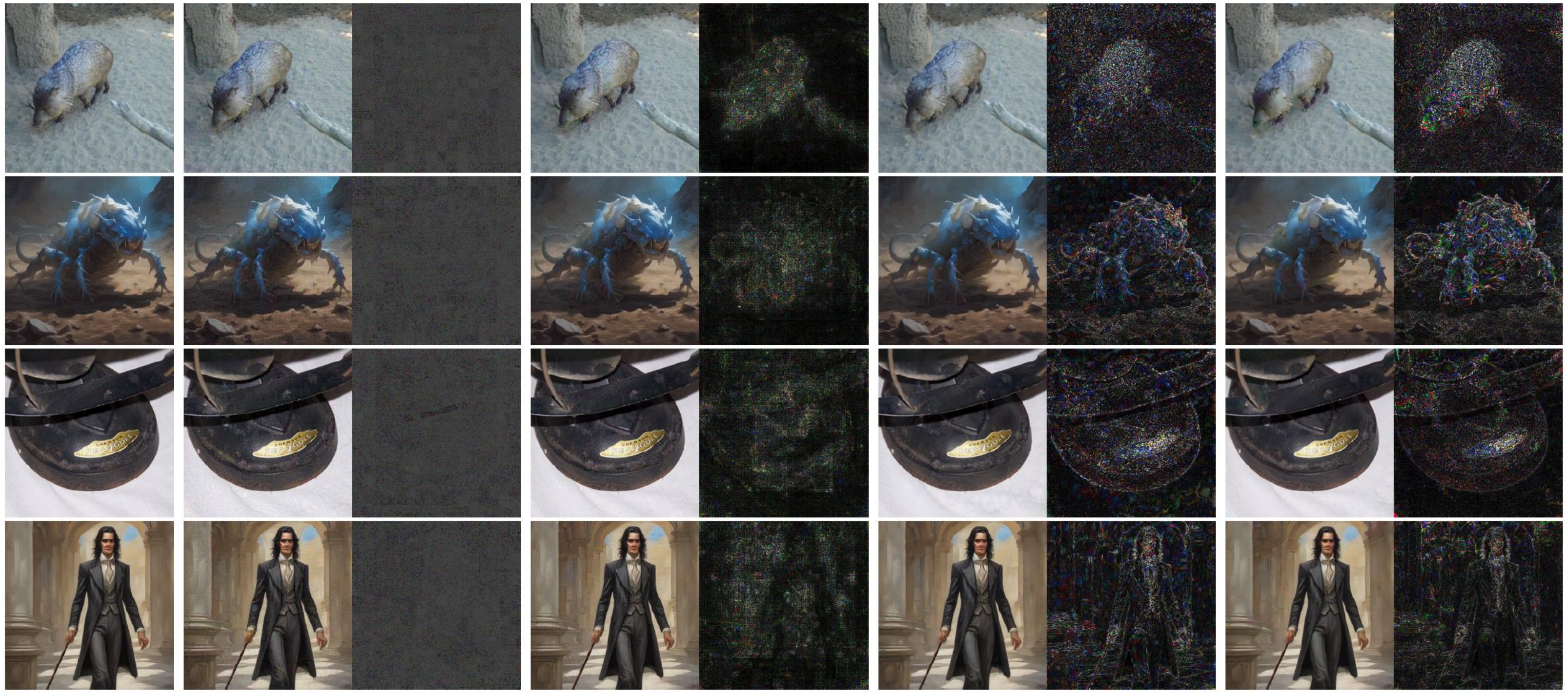
$$\mathcal{L} = \mathcal{L}_m(I_w) + \mathcal{L}_m(I_{p1}) + \mathcal{L}_m(I_{p2}) + \lambda_p\mathcal{L}_p(I_w, I) + \lambda_i\mathcal{L}_i(I_w, I)$$

Benchmark

Method	PSNR	SSIM	Bit Accuracy									
			None	Brightness	Contrast	JPEG	Gau. blur	Gau. noise	VAE-B	VAE-C	Diffusion	Avg
ImageNet												
DwtDctSvd[14]	**39.67**	**0.978**	0.993	0.636	0.489	0.848	0.992	**0.993**	0.550	0.562	0.592	0.739
±std	1.939	0.011	0.049	0.307	0.222	0.147	0.058	0.051	0.063	0.078	0.106	N/A
SSL Watermark[20]	31.04	0.862	**1.000**	**1.000**	**1.000**	0.972	**1.000**	0.937	0.793	0.777	0.743	0.914
±std	0.110	0.029	0.000	0.000	0.000	0.034	0.000	0.028	0.073	0.096	0.077	N/A
Stable Signature[19]	28.74	0.838	0.978	0.971	0.937	0.832	0.859	0.892	0.630	0.645	0.534	0.809
±std	3.246	0.080	0.054	0.061	0.092	0.106	0.121	0.117	0.086	0.105	0.064	N/A
FreqMark(Ours)	31.27	0.857	**1.000**	0.995	**1.000**	**0.991**	**1.000**	0.939	**0.938**	**0.924**	**0.969**	**0.973**
±std	3.359	0.038	0.000	0.028	0.000	0.024	0.000	0.088	0.083	0.081	0.052	N/A
DiffusionDB												
DwtDctSvd[14]	**39.49**	**0.978**	**1.000**	0.607	0.457	0.887	**1.000**	**1.000**	0.563	0.556	0.569	0.738
±std	1.182	0.006	0.000	0.308	0.194	0.109	0.000	0.000	0.053	0.059	0.085	N/A
SSL Watermark[20]	31.01	0.827	**1.000**	**1.000**	**1.000**	0.956	**1.000**	0.954	0.742	0.744	0.729	0.903
±std	0.064	0.027	0.000	0.000	0.000	0.048	0.000	0.037	0.109	0.102	0.081	N/A
Stable Signature[19]	28.31	0.844	0.996	0.996	0.990	0.896	0.858	0.967	0.668	0.733	0.527	0.848
±std	1.608	0.033	0.013	0.012	0.014	0.042	0.086	0.028	0.063	0.049	0.040	N/A
FreqMark(Ours)	31.20	0.854	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	0.934	**0.925**	**0.897**	**0.945**	**0.967**
±std	1.538	0.029	0.000	0.000	0.000	0.000	0.000	0.061	0.066	0.059	0.047	N/A

Results

Why the Image Latent Frequency Space?



Origin | Pixel | Pixel Frequency | Latent | Latent Frequency

Location	PSNR	SSIM	Bit Accuracy									
			None	Brightness	Contrast	JPEG	Gau. blur	Gau. noise	VAE-B	VAE-C	Diffusion	Avg
Pixel	**31.36**	0.771	0.950	0.935	0.937	0.848	0.885	0.925	0.642	0.654	0.542	0.813
Pixel Frequency	31.31	0.809	**1.000**	**1.000**	**1.000**	0.950	0.937	**1.000**	0.797	0.775	0.596	0.895
Latent	31.35	**0.886**	0.994	0.993	0.981	0.906	0.979	0.804	0.796	0.833	0.675	0.885
Latent Frequency	31.20	0.854	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	0.934	**0.925**	**0.897**	**0.945**	**0.967**

NEURAL INFORMATION PROCESSING SYSTEMS | 2024

FreqMark: Invisible Image Watermarking via Frequency Based Optimization in Latent Space

Thanks!

For more information, please refer to our full paper published in NeurIPS 2024!