SUPPLEMENTARY MATERIAL

The supplementary material provides a detailed explanation of the noise layer and the discriminator outlined in the letter, along with additional experimental results.

A. Noise Layer

To improve the robustness of the watermarking framework against various channel attacks, the noise layer comprised of different simulated attacks is developed, including JPEG (implemented by Forward ASL [26] at QF=30), Gaussian blur with various kernels (σ=2, $k \in \{3, 5, 7\}$), Gaussian noise (σ=0.05), Salt and Pepper noise (p=0.05), and Median blur ($k \in \{3, 5, 7\}$). In the training phase, the noise layer randomly chooses one type of attack for each mini-batch with equal probability.

B. Discriminator

The discriminator is implemented using 4 convolutional layers (Conv+BatchNorm+ReLU), followed by one global average pooling layer, one linear layer and Sigmoid activation. In the training stage, the discriminator and the encoder are alternately optimized.

C. Illustration of Watermarked Images

To qualitatively assess the image quality of different watermarking schemes, the watermarked images are illustrated in Fig. 3. It is observed that the embedding modifications across all watermarking schemes are imperceptible due to high signal-to-noise ratio, and different schemes produce distinct watermark residual patterns, which emerge as a factor influencing the robustness.

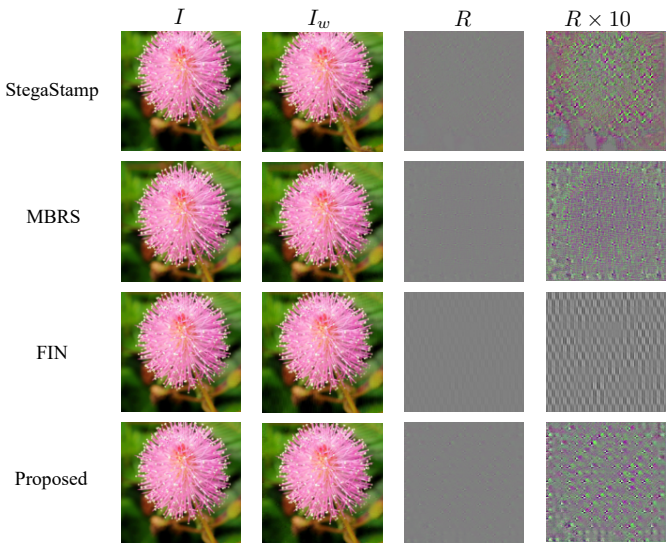

Fig. 3. The watermarked images under different schemes. The watermark residuals in the last column are enhanced tenfold to highlight the differences.

D. Expand to Images with Various Resolutions

To verify the scalability of the proposed HiFiMSFA for practical applications, we conduct further experiments in terms of image resolution. The DIV2K dataset [27] is utilized for evaluation purpose since it contains high-quality images. The pre-trained HiFiMSFA is applied directly to images of varying resolutions, with the embedding capacity adjusted accordingly, as it operates independently of the input size. The experimental results are shown in Table VI and the watermarked images are illustrated in Fig. 4. It is observed that the proposed HiFiMSFA can be effectively applied to images with higher resolutions while ensuring excellent image quality and robustness performance.

TABLE VI
DECODING ACCURACY (%) OF THE PROPOSED HIFIMSFA FOR DIFFERENT IMAGE RESOLUTIONS. "G.", "S & P", "M." DENOTE GAUSSIAN, SALT AND PEPPER, MEDIAN RESPECTIVELY.

Res.	PSNR	JPEG QF=30	G.Blur σ=2,k=7	G.Noise σ=0.05	S & P p=0.05	M.Blur k=7	Crop p=0.5
256^2	42.69	97.72	99.56	99.09	99.77	96.80	99.54
400^2	42.75	97.76	99.69	99.06	99.79	97.58	99.55
512^2	42.79	97.75	99.74	99.00	99.77	97.81	99.60
720^2	42.82	97.73	99.77	99.02	99.79	98.17	99.64

E. Illustration of OSN Effects

To demonstrate the influence of practical transmission channels, the watermarked images, distorted images, and their differences are displayed in Fig. 4. It is ready to see that the watermarked images experience noticeable distortions introduced by the WeChat channel.

Fig. 4. Illustration of the watermarked images of varying resolutions generated by HiFiMSFA and the effects of the corresponding WeChat distorted images.

F. Comparison with More SOTA Schemes

We provide further comparison experiments with the SOTA schemes, i.e., ARWGAN [28] and SEDD [29] in Table VII and Table VIII, respectively. Since the source code for SEDD is not provided, the experimental results presented in the original paper are utilized for comparison. It can be observed that the

TABLE VII
THE COMPARISON OF HIFIMSFA WITH ARWGAN [28] IN TERMS OF PAYLOAD, IMAGE QUALITY, AND DECODING ACCURACY (%) UNDER THE IMAGE SIZE OF 128×128.

Scheme	Payload	PSNR	LPIPS ↓	SSIM ↑	JPEG QF=50	GaussBlur σ=2	Crop p=0.5
ARWGAN [28]	30	37.82	0.709	0.964	88.90	95.12	99.04
Proposed	64	43.06	0.419	0.974	97.32	99.52	99.53

TABLE VIII
THE COMPARISON OF HIFIMSFA WITH SEDD [29] IN TERMS OF PAYLOAD, IMAGE QUALITY, AND DECODING ACCURACY (%) UNDER THE IMAGE SIZE OF 128×128.

Scheme	Payload	PSNR	SSIM ↑	JPEG QF=50	GaussBlur σ=2.5	SaltPepper p=0.1	MedianBlur k=7	Crop p=0.4
SEDD [29]	30	35.79	0.907	99.94	99.79	100	99.67	100
Proposed	64	36.72	0.907	99.96	99.99	99.98	99.99	99.93

proposed scheme demonstrates a general superiority across payload, image quality, and decoding accuracy.

G. Ablation Study on Loss Weights

We conduct further ablation experiments by adjusting the initial message loss weight λ_1 while keeping its decay rate and other loss weights unchanged. The results are presented in Table IX. It can be observed that as λ_1 increases, the image quality deteriorates, while the robustness improves gradually. Similarly, adjusting the image loss weight also involves a trade-off between image quality and robustness. Therefore, the loss weights can be adaptively adjusted based on specific requirements.

TABLE IX
ABLATION STUDY OF THE MESSAGE LOSS WEIGHT IN TERMS OF IMAGE QUALITY AND DECODING ACCURACY (%).

Loss weight	λ_1=3	λ_1=5	λ_1=8	λ_1=13
PSNR	44.12	43.42	41.85	41.37
LPIPS ↓	0.415	0.456	0.613	1.033
SSIM ↑	0.974	0.972	0.963	0.959
JPEG (QF=30)	94.89	95.47	97.74	98.64
GaussBlur (σ=2)	98.86	99.33	99.65	99.77
GaussNoise (σ=0.05)	97.33	97.66	99.31	99.40
SaltPepper (p=0.05)	99.54	99.41	99.67	99.88
MedianBlur (k=7)	96.26	97.31	98.56	98.29
Crop (p=0.5)	99.22	99.12	99.80	99.79

REFERENCES

[26] Zhang, Chaoning and Karjauv, Adil and Benz, Philipp and Kweon, In So, "Towards robust deep hiding under non-differentiable distortions for practical blind watermarking," in *Proc. ACM MM*, 2021, pp. 5158–5166.

[27] Agustsson, Eirikur and Radu Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," in *Proc. CVPRW*, 2017, pp. 126–135.

[28] Huang, Jiangtao and Luo, Ting and Li, Li and Yang, Gaobo and Xu, Haiyong and Chang, Chin-Chen, "ARWGAN: Attention-guided robust image watermarking model based on GAN," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–17, 2023.

[29] Xiang, Yuyuan and Wang, Hongxia and Yang, Ling and He, Mingze and Zhang, Fei, "SEDD: Robust Blind Image Watermarking With Single Encoder And Dual Decoders," *The Computer Journal*, vol. 67, no. 6, pp. 2390-2402, 2024.