SUPPLEMENTARY MATERIAL

This supplementary material first summarizes our contributions, and then elaborates on the design of the wavelet decoder, the determination of the loss weights, and the results from extended experiments using different display and shooting devices, as well as various shooting angles. Finally, we include the practical tests on print-scanning and print-camera, and compare the complexity among the involved schemes. The following sections provide detailed insights into these aspects.

A. Summary of our contributions

- **Wavelet-Domain Embedding**. We incorporate the wavelet integrated CNNs and propose a Deep Wavelet-domain Watermarking (DWW) model, dedicated to embedding watermarks in the wavelet domain rather than the spatial domain utilized by the previous arts.
- **Enhanced Loss Function**. We develop a frequency-domain enhanced mask loss to increase the loss weights in the image high-frequency regions of the image, encouraging the model to embed information into the low-frequency components in priority, thereby improving robust performance.
- Experiment results show that the proposed DWW consistently outperforms the SOTA schemes by a clear margin in terms of embedding capacity, imperceptibility, and robustness.

B. Design of the decoder

The wavelet decoder shown in Fig. 1 is designed to efficiently extract messages from watermarked images that have undergone attacks. It comprises two principal functions: multiscale feature extraction and message retrieval. For the extraction of features, the decoder employs three specifically tailored DWT blocks. Each of these blocks decomposes the input feature maps into four distinct wavelet sub-bands: LL, LH, HL, and HH. In a DWT block, the LL sub-band is further processed through a convolution block and then fed into the next DWT block for higher-level feature extraction. The other high-frequency sub-bands (LH, HL, HH) are decomposed obtain the LL sub-bands. This approach ensures the specialized feature extraction for different wavelet components. It is crucial to note that the decoder's input is in the form of wavelet sub-bands.

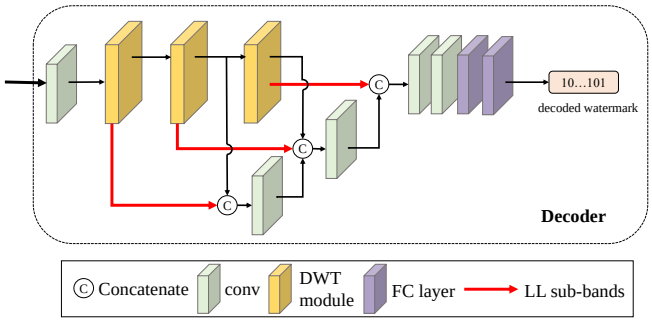

Fig. 1. The structure of wavelet decoder.

C. Structure of the DWT/IDWT module in the encoder and the DWT module in the decoder

The DWT and IDWT modules in the encoder are shown in Fig. 2, and the DWT module in the decoder is illustrated in Fig. 3. For specified details, please refer to the original WCNN.

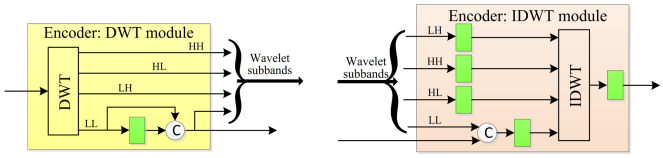

Fig. 2. The structure of DWT module and IDWT module in the encoder.

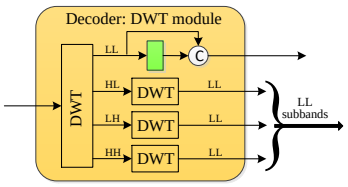

Fig. 3. The structure of DWT module in the decoder.

D. The determination of the loss weights

Specifically, λ_{spa} is set to 2, the λ_l is set to 1.5, λ_d is maintained at 0.5, λ_s is set to 3, and λ_{fre} is set to 5. These values are determined based on extensive experimentation to ensure optimal model performance. It should be noted that except for λ_s, all the other parameters are dynamic, which linearly increase from 0 to the preset values in the first 1500 iterations.

E. Different display and shooting devices

To test the adaptability of the proposed scheme to different display devices and shooting devices in real-world scenarios, we conduct additional screen-shooting experiments on different types of screens and cell phones. The shooting distance is fixed as 50cm along with the shooting angle of 0°. The experimental results are shown in Table I, where the decoding accuracy of the proposed scheme consistently exceeds 98%, indicating its adaptability for different devices.

F. Different shooting angles

In order to test the screen-shooting robustness at different shooting angles, we display the watermarked images on the screen and use the mobile phone to shoot the images at different angles. The experiments are carried out with the display device "PHILPS 275E9" and the shooting device "iPhone 11". The shooting distance is fixed as 30cm, and the shooting angle ranges from left 40° to right 40° with a step of 10°. Fig. 4 illustrates the captured images with different shooting angles (ranged from left 10° to left 40°), and it is observed that as the shooting angle increases, the perspective distortion become more serious. The experimental results are

shown in Table II and Table III. We can see that DWW exhibits remarkable decoding accuracy above 97% for both 100-bit and 300-bit capacities, and also exhibits consistent robustness at a wide range of shooting angles. Additionally, the image quality of DWW outperforms StegaStamp, WCNN, and PIMoG.

TABLE I
DECODING ACCURACY WITH DIFFERENT DEVICES.

Display devices	Shooting devices		
	iPhone 11	vivo Z5X	iPhone 13
HKC MG27Q	99.14	98.36	99.35
PHILIPS 275E9	98.97	99.06	99.11

TABLE II
DECODING ACCURACY IN REAL-WORLD SCENARIOS (L→LEFT).

Scheme	Capacity (bits)	Image size	PSNR	Shooting angles			
				L-40°	L-30°	L-20°	L-10°
StegaStamp	100	400	30.28	**98.40**	98.46	98.74	98.91
WCNN	100	400	33.30	97.00	97.99	97.26	97.95
DWW	100	400	**39.61**	98.31	**98.97**	**99.07**	**99.25**
PIMoG	30	128	36.42	95.33	**97.58**	96.44	97.01
DWW	300	400	**37.11**	**97.39**	97.31	**97.18**	**97.52**

TABLE III
DECODING ACCURACY IN REAL-WORLD SCENARIOS (R→RIGHT)

Scheme	Capacity (bits)	Image size	PSNR	Shooting angles			
				R-10°	R-20°	R-30°	R-40°
StegaStamp	100	400	30.28	98.01	97.14	96.36	97.73
WCNN	100	400	33.30	98.02	97.65	97.51	96.71
DWW	100	400	**39.61**	**99.34**	**98.86**	**98.53**	**97.93**
PIMoG	30	128	36.42	98.07	95.67	92.57	93.79
DWW	300	400	**37.11**	**98.29**	**97.94**	**97.81**	**97.72**

Left 40° Left 30° Left 20° Left 10°

Fig. 4. Example of screen-shooting images at different shooting angles.

G. Practical test on print-scaning and print-camera

While the practical test results in the manuscript only demonstrate that our proposed DWW provides excellent resistance to screen-shooting attacks, it is imperative to clarify that its utility is not confined to this scenario alone. It also excels in print-scaning and print-camera tests, establishing its efficacy across diverse real-world applications. The performance outcomes for these scenarios are depicted in Table IV. These results corroborate the algorithm's versatility.

TABLE IV
PRINT-SCAN AND PRINT-PHONE SCENARIOS WITH THE SOTA SCHEMES.

Schme	print-scaning	print-camera
StageStamp	97.69	98.26
WCNN	97.91	95.78
PIMOG	96.52	93.37
DWW	**98.50**	**98.73**

H. Complexity comparison

In Table V, we present a thorough comparison of DWW against current state-of-the-art (SOTA) schemes, evaluated in terms of model parameters and floating-point operations (FLOPs). The results highlight DWW's high efficiency in that it outperforms other SOTA schemes with fewer parameters and computational overhead.

TABLE V
COMPLEXITY COMPARISON OF OUR DWW IN TERMS OF MODEL PARAMETERS (PARAMS.) AND FLOATING-POINT OPERATIONS (FLOPS) WITH THE SOTA SCHEMES.

Method	Params. (↓)	FLOPs (↓)
StageStamp	52.37M	110M
WCNN	33.22M	73M
DWW	**32.20M**	**71M**