Robust Image Watermarking Scheme under Halftone Distortion with Surrogate Model

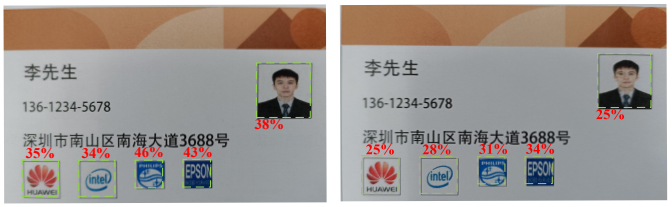
Changsheng Chen* and Xijin Li
Guangdong Provincial Key Laboratory of Intelligent Information Processing, Shenzhen University, Shenzhen, China
Shenzhen Key Laboratory of Media Security, Shenzhen University, Shenzhen, China
*E-mail: cschen@szu.edu.cn

***Abstract*—Recently, document images have been widely used in various online applications. Digital watermarking is an important forensic technique to verify the authenticity of a document image. However, the recapturing operation leads to a significant risk in document images since the extraction accuracy of recaptured digital watermarking drops significantly. In this work, we propose robust watermark against halftone distortion by utilizing surrogate models and end-to-end watermarking frameworks. Firstly, we employ differentiable surrogate models to generate the halftone distortion. Then, surrogate models are incorporated into end-to-end watermarking frameworks to enhance the robustness of the watermark in the print-camera scenario. To evaluate the robustness of the proposed method, we conduct a series of experiments in real-world scenarios. The experimental results confirm that our method can improve the robustness of the watermark across different devices, datasets, printing sizes, and watermark capacities.**

I. Introduction

The document image is an important information carrier widely used in e-business and e-government applications. However, due to the development of image editing tools and image generation methods, document image security is facing great threats. Active image forensic techniques like digital watermarking and passive image forensic techniques like tamper detection are commonly utilized to ensure the security of document images. However, researchers have found that recapturing operations significantly reduces the effectiveness of existing document image forensic methods. For instance, Zhao *et al* [1]. utilize a deep learning scheme to edit practical document images and conceal the traces of tampering by recapturing operations. Moreover, as shown in Fig. 1, recaptured watermarked images suffer severe distortions, resulting in a significant decline in extraction accuracy. Specifically, the error bit rate of watermarks in small printing sizes is close to 50%. Therefore, research on document image forensics techniques in recapturing scenarios is of practical significance and demands our attention. In this work, we focus on improving the robustness of digital watermarking in the print-camera scenario.

Previous research in digital watermarking has mainly focused on robustness against digital channel distortion, which mainly consists of Gaussian noise, Gaussian blur, and other factors. Recently, capturing a printed image with a mobile phone has gained popularity as information acquisition. Compared to digital channel distortion, print-camera channel distortion is much more complex, including perspective distortion, color distortion, and environmental noise, which brings significant challenges to digital watermarking. Therefore, robust watermark in the print-camera scenario has gained great attention in recent years. However, most existing works focus on the distortions introduced by the shooting process while neglecting halftone distortion. Through our research on the distortion produced by the print-camera process, we find that halftoning technology is widely used in current mainstream printers, which means that the halftone distortion is a key part of the print-camera distortion.



(a) StegaStamp (b) StegaStamp+S

Fig. 1. The practical application of robust watermark against halftone distortion. The images are embedded with 200 bits with printing sizes of 1.5×1.5 cm^2 and 0.75×0.75 cm^2. The bit error rate is shown next to the corresponding watermark image. (a) The application example of StegaStamp [2]. (b) The application example of our method.

Given the above limitations, we propose robust watermark against halftone distortion. This approach aims to enhance the robustness of digital watermarking and improve the security of current image forensic systems. We train a CycleGAN-based simulation network for the halftone distortion and employ end-to-end watermarking frameworks based on deep learning. To evaluate the robustness of the proposed method, we conducted a series of experiments in real-world application scenarios, including different devices, datasets, printing sizes, and watermark capacities. The experimental results demonstrate that the proposed method achieves better performance than SOTA methods, and the average bit error is lower than 20% in most scenarios.

The main contributions of this work are as follows:

1) We employ a simulation network of the halftone distortion based on CycleGAN to enhance the robustness of the watermark. Moreover, this simulation network can be incorporated into different deep learning-based watermarking frameworks.

2) Experimental results confirm that our method achieves better robustness in several real-world scenarios, including different devices, datasets, printing sizes, and watermark ca-

pacities.

II. The Proposed Method

A. Surrogate Model for Halftone Distortion

1) The Printing Process (Inkjet, LaserJet): The halftoning technique is utilized by printers to simulate a color image in continuous tone through halftone dots with variable sizes or spacing. Dispersed dot and cluster dot halftoning [3] are two classes of halftoning patterns that are widely used by inkjet printers and laserjet printers, respectively.

Mainstream inkjet printers employ the error diffusion halftoning technique to produce the dispersed dot halftoning pattern. During the error diffusion process, each color channel is processed with the following steps [3]. First, each pixel value of the input image is set to 0 or 255 according to the threshold value of 128. Second, the error value is calculated between the assigned value in the first step and the actual value. Finally, according to a predefined error filter kernel, the error is accumulated and diffused to the neighborhood pixels. In our implementation, we adopt the error filtering kernel defined by Floyd and Steinberg [4], and weights in this filter are [0, 0, 7/16; 3/16, 5/16, 1/16].

Mainstream laserjet printers produce the cluster dot. When we consider a single-channel image patch, the representation of the halftone can be described as follows

$$I_L(\mathbf{x}) = \sum_m \sum_n \delta(\mathbf{x} - m\mathbf{a} - n\mathbf{b}) \otimes H(\mathbf{x}), \tag{1}$$

where $\mathbf{x} = (x, y)$ represents the horizontal and vertical coordinates of each pixel of the image, 'δ' denotes the Dirac delta function, spatial halftone vectors **a** and **b** are the density and direction of the halftone array, m, n represent indices of halftone dots along the direction of **a** and **b**, respectively, '$\otimes$' denotes the convolution operation, and $H(\mathbf{x})$ is a binary masking function that defines the shape of halftone dots. In our implementation, we adopt the ordered dithering technique with a halftone dot size of 4×4 pixels. When considering a color image, the halftoning process described in Eq. (1) operates independently in each color channel.

2) A Surrogate Model for Halftone Distoriton: The noise layer is vital in our method. Moreover, the end-to-end watermarking framework requires a differentiable noise layer. Chen *et al*. [5] propose a distortion model guided surrogate model based on CycleGAN structure with two training stages to generate distortions introduced by the recapturing process. The trained surrogate model can be utilized to improve the generalization performance of networks. What's more, the surrogate model consists of a series of differentiable operations, which meets the requirement of the end-to-end watermarking framework.

Therefore, inspired by [5], we utilize CycleGAN-based surrogate models for generating halftone distortions, aiming to improve the robustness of the watermark. We collect three types of images for the surrogate model training, including original images I_O, simulated halftone images I'_H yield by the process described in Sec. II-A1, and real halftone images I_H. The training process of the surrogate model is shown in Fig. 2. In the first stage, I_O and I'_H are employed to train the surrogate model for the basic ability to generate halftone distortions. In the second stage, real halftone images I_H are utilized to finetune the discriminator. In addition, I_O and I'_H are still participating in training to provide a reliable source for learning halftone distortion.

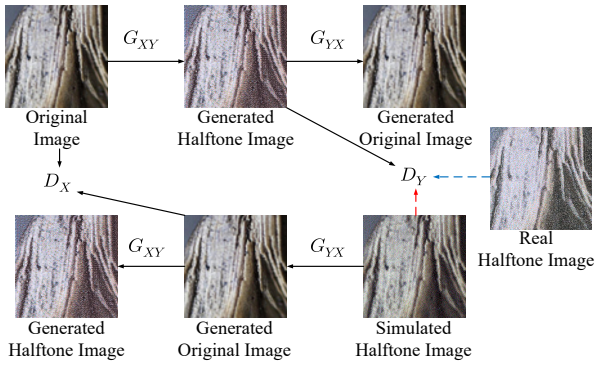


Fig. 2. The surrogate model for halftone distortion. The black arrows indicate the steps for two stages. The red dashed line represents the step only for the first stage. The blue dashed line represents the step only for the second stage.

B. Watermarking Framework

We adopt the end-to-end watermarking framework to realize robust watermark against halftone distortion, which consists of an encoder, noise layer, decoder, and discriminator. Specifically, we improve the noise layer by utilizing surrogate models for halftone distortion. The generic watermarking framework is shown in Fig. 3. The UNet-like encoder embeds the watermark and maintains the quality of the watermarked image. The encoder converted the concatenated watermark and original image to a residual image, which is overlaid on the original image to finish the watermark embedding. The decoder consisting of convolution layers and fully connected layers recovers the watermark from the distorted image by converting the image to a vector with the same length as the original watermark. The discriminator outputs a binary sequence to judge whether the image is watermarked. Similar to generative adversarial networks, adversarial training between the discriminator and encoder can further improve the visual quality of the watermarked image. The overall loss function of the generic watermarking framework can be written as

$$\mathcal{L} = \lambda_1 \cdot \mathcal{L}_I + \lambda_2 \cdot \mathcal{L}_W + \lambda_3 \cdot \mathcal{L}_D, \tag{2}$$

where $\mathcal{L}_I$, $\mathcal{L}_W$, and $\mathcal{L}_D$ denote the image reconstruction loss, watermark loss, and discriminator loss. λ_1, λ_2, λ_3 are the weights for three loss components. Image reconstruction loss can be MSE loss, LPIPS perceptual loss [6], etc, which are utilized to improve the visual quality of the watermarked image. Discriminator loss aims to distinguish the original image and the watermarked image and usually adopts Wasserstein loss [7], which is commonly utilized in generative adversarial

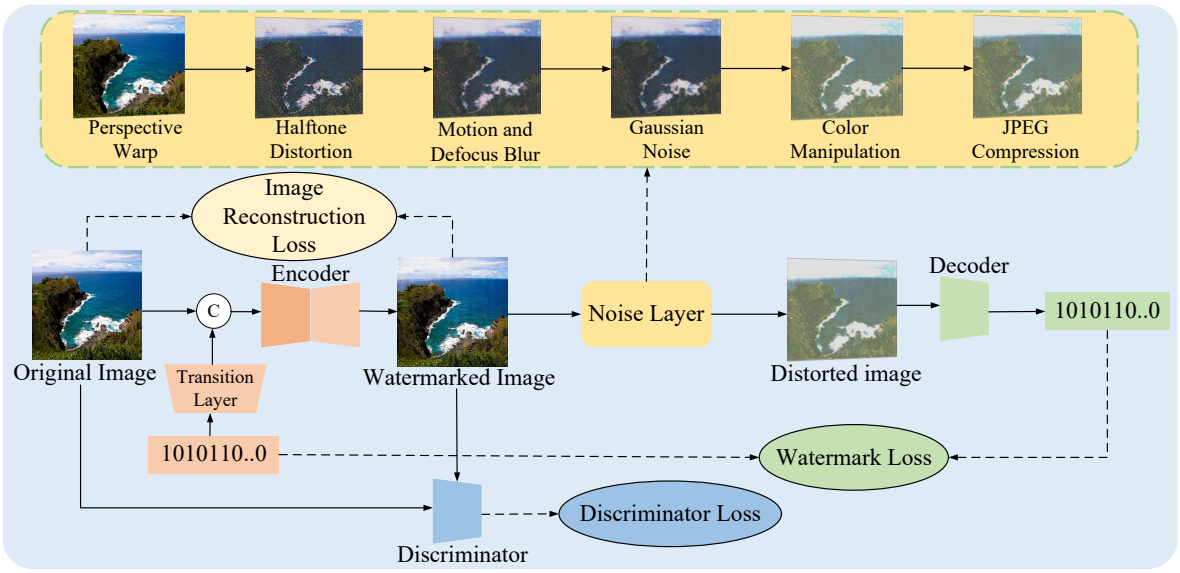


Fig. 3. The framework of robust watermark against halftone distortion.

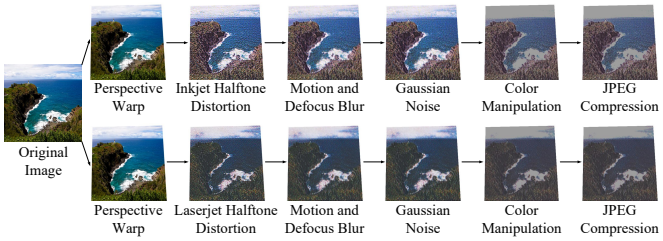


Fig. 4. The visualization of each step of distortion layers.

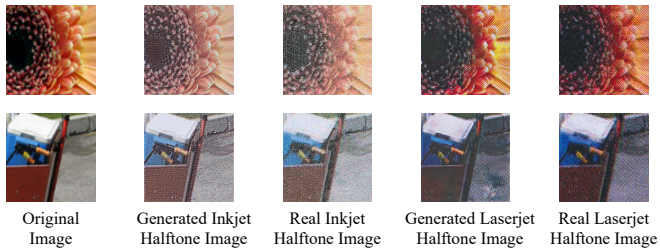


Fig. 5. Halftone images generated by surrogate models and collected in the real environment.

networks. Watermark loss aims to minimize the difference between the original watermark and the recovered watermark, utilizing cross entropy loss.

The noise layer is the key to robust watermark against halftone distortion. The noise layer in StegaStamp mainly focuses on the distortion introduced by the shooting process, while ignoring the halftone distortion in the printing process. Our surrogate model focuses on generating the halftone pattern during the printing process. Therefore, we combine the halftone distortion surrogate model with the noise layer proposed by StegaStamp. The visualization of each step of distortion layers is shown in Fig. 4. The details of the noise layer are shown as follows:

1) Perspective Warp: The unevenness of the paper during the printing process and the deviation of the lens during the shooting process can cause perspective deformation. To simulate this distortion, a random perturbation is employed to the corner locations of watermarked images ($\pm$ 40 pixels), and homographic transformation is adopted to map the watermarked images to new locations.

2) Halftone Distortion: Two surrogate models are trained to simulate inkjet and laserjet halftone distortion, respectively. As shown in Fig. 5, surrogate models can generate halftone distortions similar to real halftone images. To avoid the watermark network overfit halftone distortion, both of them are added to the noise layer and the probability is both set as 50%. The halftone distortion enables the encoder to adjust the embedding area, intensity, etc automatically. At the same time, it also improves the decoder's ability to recover messages from watermarked images with halftone distortions.

3) Motion and Defocus Blur: The shooting process can introduce blur because of the camera motion and inaccurate autofocus. A random angle between 0 and 2π and a straight line blur kernel with a width between 3 and 7 pixels are employed to simulate motion blur. In addition, a Gaussian blur kernel with a standard deviation sampling between 1 and 3 pixels is adopted to simulate defocus blur.

4) Gaussian Noise: The electronic components in the camera will introduce various noises during the imaging process. A Gaussian noise with the standard deviation sampling in [0, 0.02] is employed to simulate the noise.

5) Color Manipulation: The camera uses white balance, exposure settings, and chroma correction to adjust the output image. These distortions can be approximated with random color transformations, including hue shift, desaturation, and brightness adjustment. the hue shift can be achieved by adding a random color offset sampled uniformly from [-0.1, 0.1] to each RGB channel. The desaturation can be achieved by randomly linearly interpolating between the full RGB image

and its grayscale equivalent. The brightness adjustment can be achieved by a linear transformation $mx + b$ with m in [0.5, 1.5] and b in [-0.3, 0.3].

6) JPEG Compression: The images captured by the camera are usually saved in the JPEG format. Since the rounding step of JPEG compression is not differentiable at zero, the method proposed by Shin *et al* [8] is employed to approximate the quantization step near zero with a piecewise function. The JPEG quality is sampled uniformly in [50, 100].

III. Experimental Results

To train the surrogate model for halftone distortion, we collect the training data with two printers and a scanner. We choose an inkjet printer (Canon G3800) and a laser printer (Canon iR-ADV C3530), which are two main types of printers. Considering that the camera-shooting process may introduce various factors, such as lighting and noise, we utilize a high-quality scanner (CanoScan 5600F) to capture the halftone distortion of the printed images minimizing the factors that affect the halftone distortion capturing as possible. First, We randomly select 300 images from the MIRFLICKR dataset [9] and print them with the size of 5 × 5 cm^2. Then, we scan the printed images and crop them to 400 × 400 resolution. Finally, we collect 15000 patches with halftone distortions for the training of surrogate models.

To train the watermarking network against halftone distortion, we randomly select 20,000 images from the MIRFLICKR dataset [9] as the training dataset. The training images are resampled to 400 × 400 resolution and are embedded with 200 bits random message. The implementation is based on Tensorflow 1.13.1.

We evaluate the robustness performance of the proposed method in the real-world environment. Firstly, 20 images not included in the training set are randomly selected from the MIRFLICKR dataset for watermark embedding. The length of the watermark is 200 bits. Secondly, the watermarked images are printed by inkjet and laserjet with the printing size of 2 × 2 cm^2 on A4 standard paper. Thirdly, the printed images are captured by mobile phones with the shooting distance and angle of 10 cm and 0°, respectively. Finally, the decoder recovers messages from captured images.

The bit error rate (BER) is employed to evaluate the robustness of different methods. A lower bit error rate indicates better robustness. Two SOTA print-camera resistant watermarking methods, StegaStamp [2] and NSN [10], are used for comparison. Our surrogate models are incorporated into these SOTA methods. The details of our experimental configurations are shown below.

StegaStamp: This approach proposes a set of differentiable operations to simulate distortions in the print-camera scenario, including perspective warp, motion and defocus blur, color manipulation, noise, and JPEG compression.

NSN: This approach proposes a noise simulation network. In the earlier training stage, the math model is utilized to simulate distortions in the print-camera scenario. In the later training stage, the math model is replaced with the noise simulation network.

StegaStamp+S: Our noise layer consists of **s**urrogate (S) models for the halftone distortion and differentiable operations proposed by StegaStamp, which generate printing distortion and shooting distortion, respectively. The other configurations (encoder, decoder, training loss, and hyperparameters) of *StegaStamp+S* are the same as those of *StegaStamp*.

NSN+S: In the later training stage, the noise simulation network is replaced with our noise layer which contains surrogate models for halftone distortion. The other configurations (encoder, decoder, training loss, and hyperparameters) of *NSN+S* are the same as those of *NSN*.

To validate the effectiveness of the proposed method, we conduct a series of robustness experiments in different real-world scenarios, including different devices, datasets, printing sizes, and watermark capacities. Detailed experimental results are shown as follows.

A. Visual Quality Evaluation

In this part, we evaluate the visual quality of watermarked images after printing and shooting both quantitatively and qualitatively.

The quantitative comparisons are shown in Tab. I. We employ Peak Signal Noise Ratio (PSNR), Structural Similarity (SSIM) [11], and Frechet Inception Distance (FID) [12] to evaluate the visual quality of watermarked images. It can be seen that watermarked images after the print-camera process perform poorly in visual quality performance indicators. This is because watermarked images undergo various distortions, including halftone distortion, noise, blurring, etc. Moreover, *StegaStamp* and *NSN* achieve slightly better performance than *StegaStamp+S* and *NSN+S* in visual quality. This is because the surrogate model generates obvious halftone distortion, which leads to severe damage to the watermark. Therefore, the encoder continuously adjusts the embedding area and intensity to ensure correct watermark decoding after halftone distortion while sacrificing some visual quality.

The qualitative comparisons are shown in Fig. 6. It can be seen that watermarked images of *StegaStamp+S* and *NSN+S* contain foggy traces visually. This is because the encoder adjusts embedding strength and area of the watermark to combat halftone distortion.

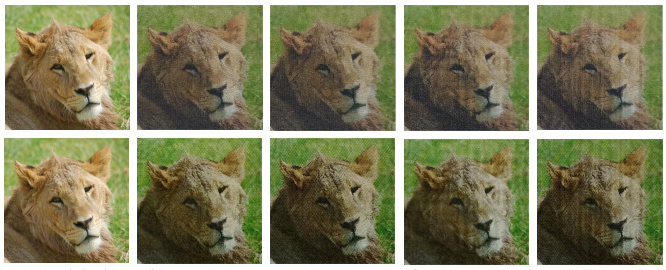
The watermarked images in practical scenarios are usually printed in a small size. Moreover, they undergo various distortions after printing and shooting. The slight decrease in visual quality after using the distortion surrogate model has a relatively small impact and is within the acceptable range. Therefore, we mainly focus on the robustness of the watermark in practical scenarios.

B. Robustness Evaluation

1) Robustness to Different Printing and Shooting Devices: In practical scenarios, watermarked images are commonly captured by various printers and mobile phones with different brands and imaging qualities. Therefore, it is essential to

TABLE I
THE VISUAL QUALITY OF DIFFERENT METHODS. THE BEST PERFORMANCES ARE **BOLD-FACED**.

Methods	Metrics		
	PSNR(dB)	SSIM	FID
StegaStamp	**14.14**	0.2435	**262.26**
NSN	**14.20**	0.2514	**266.02**
StegaStamp + S	14.05	**0.2501**	265.82
NSN + S	14.03	**0.2554**	276.80



(a) Original (b) StegaStamp (c) NSN (d) StegaStamp+S (e) NSN+S

Fig. 6. The visualization of watermarked images after printing and shooting. Top: the images printed with inkjet; Bottom: the images printed with laserjet.

evaluate the robustness performance of watermarks across different devices. In this part, the watermarked images are printed by 4 printers, including inkjet (EPSON L805 and HP OfficeJet 258) and laserjet (Konica Minolta C6500 and HP LaserJet M176n). Then they are captured by 4 mobile phones with different pixel resolutions (100 MP, 64 MP, and 13 MP).

Experimental results are shown in Tab. II. It can be seen that the proposed method has achieved better robustness across different devices. The average bit error rates of our method are close to 10%. This indicates that the surrogate model effectively improves the robustness of the watermark to halftone distortion, although different printers have different parameters for the halftoning process.

2) Robustness to Different Datasets: In practical scenarios, watermarks are commonly embedded in various types of images. Identification photos and logos are two typical images widely used in important documents and are usually the key to verifying the authenticity of document images. For example, Identification photos and logos are used to identify personal and legal identities, respectively, and are high-risk areas for tampering. Therefore, we evaluate the robustness performance of watermarks across different datasets. We randomly select 20 images from the Identification Photo dataset [13] and Logo-2k+ [14], respectively. In this part, "EPSON L805" and "HP LaserJet M176n" are utilized for printing, and "Oppo K9x" is utilized for shooting.

Experimental results are shown in Tab. III. It can be observed that the proposed method has achieved better performance of each dataset with bit error rates dropping below 20%. This demonstrates the robustness of our method across different datasets, even if the content of testing images is significantly different from training images.

3) Robustness to Different Printing Sizes: In practical scenarios, watermarked images are commonly printed in different

TABLE II
BIT ERROR RATE (BER) WITH DIFFERENT PRINTING AND SHOOTING DEVICES. THE BEST PERFORMANCES ARE **BOLD-FACED**.

Devices	Methods	Honor 50se	Oppo K9x	iQoo Z5	Meizu Metal
EPSON L805	StegaStamp	26.25%	27.42%	27.55%	28.12%
	NSN	28.85%	28.55%	28.80%	29.77%
	StegaStamp + S	**11.15%**	**11.42%**	**11.67%**	**11.45%**
	NSN + S	**11.07%**	**11.22%**	**11.20%**	**12.13%**
HP OfficeJet 258	StegaStamp	22.60%	23.90%	22.87%	22.50%
	NSN	25.20%	24.55%	25.43%	24.87%
	StegaStamp + S	**9.40%**	**9.05%**	**9.25%**	**9.77%**
	NSN + S	**9.77%**	**10.28%**	**10.00%**	**11.07%**
Konica Minolta C6500	StegaStamp	21.95%	23.05%	21.52%	22.00%
	NSN	23.25%	22.30%	23.10%	22.82%
	StegaStamp + S	**10.05%**	**10.77%**	**10.50%**	**10.47%**
	NSN + S	**11.42%**	**10.82%**	**11.30%**	**11.35%**
HP LaserJet M176n	StegaStamp	25.80%	26.02%	26.80%	25.02%
	NSN	25.35%	25.15%	24.90%	24.98%
	StegaStamp + S	**10.58%**	**11.10%**	**12.17%**	**11.27%**
	NSN + S	**11.23%**	**11.55%**	**11.55%**	**11.20%**

TABLE III
BIT ERROR RATE (BER) WITH DIFFERENT DATASETS. THE BEST PERFORMANCES ARE **BOLD-FACED**.

Devices	Methods	Datasets	
		Photo	Logo
EPSON L805	StegaStamp	26.12%	26.32%
	NSN	28.78%	29.10%
	StegaStamp + S	**15.15%**	**18.07%**
	NSN + S	**13.22%**	**16.20%**
HP LaserJet M176n	StegaStamp	22.62%	22.77%
	NSN	22.02%	22.32%
	StegaStamp + S	**13.92%**	**15.77%**
	NSN + S	**11.52%**	**13.27%**

sizes according to specific requirements. For example, there is limited available space for embedding information on a personal business card, while multiple watermarked images need to be included. Therefore, it is necessary to resize watermarked images according to the importance of information or minimize the size of watermarked images as much as possible. In this part, we evaluate the robustness performance of our method across different printing sizes. The watermarked images are printed with 2×2 cm^2, 1.5×1.5 cm^2, and 1×1 cm^2. "EPSON L805" and "HP LaserJet M176n" are utilized for printing, and "Oppo K9x" is utilized for shooting.

Experimental results are shown in Tab. IV. It can be seen that the proposed method can improve the watermark robustness across different printing sizes. Specifically, our method can maintain bit error rates below 20% and 30% in printing sizes of 1.5×1.5 cm^2, and 1×1 cm^2, respectively. This indicates that the proposed method is suitable for the scenario with small and multiple watermarks. Therefore, our method enables a single image to accommodate watermarked images with different sizes, which means more information can be embedded.

4) Robustness to Different Watermark Capacities: In this part, we evaluate the robustness performance of the proposed method to different watermark capacities. We select the watermark capacities of 50 bits, 100 bits, 150 bits, and 200 bits for retraining and testing the network. "EPSON L805" and "HP LaserJet M176n" are utilized for printing, and "Oppo K9x" is utilized for shooting.

Experimental results are shown in Tab. V. It can be observed that our method can achieve better performance across differ-

TABLE IV
BIT ERROR RATE (BER) WITH DIFFERENT PRINTING SIZES. THE BEST PERFORMANCES ARE **BOLD-FACED**.

Devices	Methods	Printing size (cm^2)		
		2 × 2	1.5 × 1.5	1 × 1
EPSON L805	StegaStamp	27.42%	36.35%	43.55%
	NSN	28.55%	35.58%	42.40%
	StegaStamp + S	**11.42%**	**16.15%**	**27.65%**
	NSN + S	**11.22%**	**15.57%**	**25.48%**
HP LaserJet M176n	StegaStamp	26.02%	31.47%	40.10%
	NSN	25.15%	31.31%	41.25%
	StegaStamp + S	**11.10%**	**14.95%**	**24.37%**
	NSN + S	**11.55%**	**15.15%**	**22.85%**

TABLE V
BIT ERROR RATE (BER) WITH DIFFERENT WATERMARK CAPACITIES. THE BEST PERFORMANCES ARE **BOLD-FACED**.

Devices	Methods	Capacity (bits)			
		50	100	150	200
EPSON L805	StegaStamp	14.50%	20.30%	25.97%	27.42%
	NSN	13.50%	20.85%	22.43%	28.60%
	StegaStamp + S	**4.30%**	**11.15%**	**11.20%**	**11.42%**
	NSN + S	**2.90%**	**3.75%**	**4.57%**	**11.22%**
HP LaserJet M176n	StegaStamp	12.90%	19.40%	23.43%	26.02%
	NSN	11.50%	19.00%	20.97%	25.15%
	StegaStamp + S	**2.50%**	**8.80%**	**11.17%**	**11.10%**
	NSN + S	**3.20%**	**3.85%**	**6.27%**	**11.55%**

ent watermark capacities. Specifically, the proposed method can decrease the bit error to below 5% when the watermark capacity is 50 bits. Therefore, the watermark capacity of 50 bits can be selected when the watermark is required to be completely recovered as possible.

IV. CONCLUSION

This work proposes robust watermark against halftone distortion, utilizing surrogate models and end-to-end watermarking frameworks. The experimental results confirm that our method can achieve good robustness performance under many practical scenarios, including different devices, datasets, and printing sizes.

In the future, we plan to extend our research to robust watermark in the screen-camera scenario. By modeling distortions of the screen-camera process, especially the moiré pattern, and training the corresponding surrogate model, we may address the challenge of digital watermarking encountered in the screen-camera scenario.

ACKNOWLEDGMENT

This work is sponsored by NSFC, China 62072313, 62371301, and Guangdong Provincial Key Laboratory (Grant 2023B1212060076).

REFERENCES

[1] L. Zhao, C. Chen, and J. Huang, "Deep learning-based forgery attack on document images," *IEEE Transactions on Image Processing*, vol. 30, pp. 7964–7979, 2021.

[2] M. Tancik, B. Mildenhall, and R. Ng, "Stegastamp: Invisible hyperlinks in physical photographs," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 2117–2126.

[3] J. M. Guo and S. Sankarasrinivasan, "Digital halftone database (dhd): A comprehensive analysis on halftone types," in *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, IEEE, 2018, pp. 1091–1099.

[4] R. W. Floyd, "An adaptive algorithm for spatial gray-scale," in *Proc. Soc. Inf. Disp.*, vol. 17, 1976, pp. 75–77.

[5] C. Chen, X. Li, B. Chen, and H. Li, "A distortion model guided adversarial surrogate for recaptured document detection," *Pattern Recognition*, p. 110 433, 2024.

[6] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 586–595.

[7] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *International conference on machine learning*, PMLR, 2017, pp. 214–223.

[8] R. Shin and D. Song, "Jpeg-resistant adversarial images," in *NIPS 2017 Workshop on Machine Learning and Computer Security*, vol. 1, 2017, p. 8.

[9] M. J. Huiskes and M. S. Lew, "The mir flickr retrieval evaluation," in *Proceedings of the 1st ACM international conference on Multimedia information retrieval*, 2008, pp. 39–43.

[10] C. Qin, X. Li, Z. Zhang, F. Li, X. Zhang, and G. Feng, "Print-camera resistant image watermarking with deep noise simulation and constrained learning," *IEEE Transactions on Multimedia*, 2023.

[11] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[12] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local nash equilibrium," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[13] C. Chen, W. Chen, X. Chen, and H. Li, "A document presentation attack detection scheme with optical flow under a flashlight," *Pattern Recognition, Submitted*, 2023.

[14] J. Wang, W. Min, S. Hou, *et al.*, "Logo-2k+: A large-scale logo dataset for scalable logo classification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, 2020, pp. 6194–6201.