Watermarking One for All: A Robust Watermarking Scheme Against Partial Image Theft

Gaozhi Liu[1] Silu Cao[1] Zhenxing Qian[1*] Xinpeng Zhang[1] Sheng Li[1] Wanli Peng[2]
[1]Fudan University [2]China Agricultural University
{gzliu24, slcao23}@m.fudan.edu.cn, {zxqian, zhangxinpeng, lisheng}@fudan.edu.cn
wlpeng@cau.edu.cn

Abstract

The proliferation of digital images on the Internet has provided unprecedented convenience, but also poses significant risks of malicious theft and misuse. Digital watermarking has long been researched as an effective tool for copyright protection. However, it often falls short when addressing ***partial image theft****, a common yet little-researched issue in practical applications. Most existing schemes typically require the entire image as input to extract watermarks. However, in practice, malicious users often steal only a portion of the image to create new content. The stolen portion can have arbitrary shape or content, being fused with a new background and may have undergone geometric transformations, making it challenging for current methods to extract correctly. To address the issues above, we propose* ***WOFA*** *(Watermarking One for All), a robust watermarking scheme against partial image theft. First of all, we define the entire process of partial image theft and construct a dataset accordingly. To gain robustness against partial image theft, we then design a comprehensive distortion layer that incorporates the process of partial image theft and several common distortions in channel. For easier network convergence, we employ a multi-level network structure on the basis of the commonly used embedder-distortion layer-extractor architecture and adopt a progressive training strategy. Abundant experiments demonstrate that our superior performance in the scenario of partial image theft, offering a more reliable solution for protecting digital images against unauthorized use in practical use.*

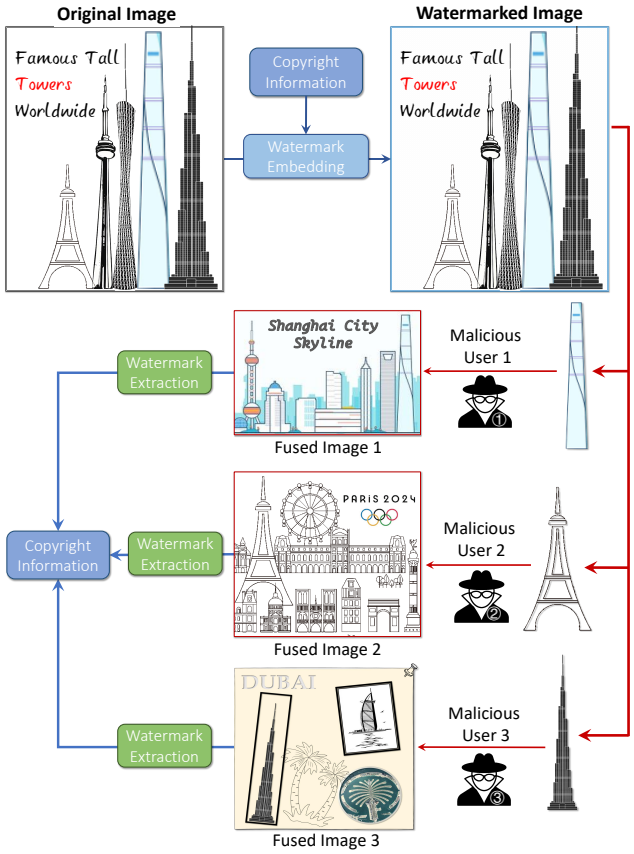

Figure 1. A brief description to the scenario of partial image theft. Malicious users may take arbitrary portion of the image to create new content.

1. Introduction

In the digital era, the widespread availability and sharing of digital images on the Internet have revolutionized how we access and disseminate information. This ease of access, however, comes with significant risks, particularly the theft and unauthorized misuse of digital content. Digital watermarking technology has been researched for long and proven to be an effective tool for copyright protection by embedding ownership information within images. However, despite its extensive research and development, existing watermarking schemes often fall short when confronted with the scenario of partial image theft.

*Corresponding author.

Partial image theft, which is a common yet little-researched issue in practical applications, occurs when malicious users steal only a portion of an original image to form new content. As shown in the Fig.1, the malicious user takes an irregularly shaped area from the original image and combines it with a new background to form a new image. From the example above, we can address that the stolen portion can be of arbitrary shape, content, composited with an unknown background and may have undergone various geometric transformations. This may bring difficulties to the existing methods in extracting the watermark. Most existing watermarking methods generally require the entire image to extract embedded watermarks, rendering them ineffective when only an irregular portion of the original image is stolen, let alone a geometric transformation might be conducted. Consequently, there is an urgent need for a watermarking solution that can robustly protect against partial image theft.

To address this gap, in this paper, we propose **WOFA** (Watermarking One for All), a novel and robust watermarking scheme designed to combat partial image theft. As the name suggests, our method only requires embedding the watermark once into the entire original image. Subsequently, the watermark can be extracted from images that contains a portion of the original image, regardless of the new image's content or the shape of the stolen portion, as shown in Fig.1. To achieve this functionality, we first define the entire process of partial image theft and create a dataset accordingly. We employ a multi-level network structure built upon the commonly used embedder-distortion layer-extractor architecture to facilitate easier network convergence. To equip the model with robustness against partial image theft, we meticulously design a progressive distortion layer that applies various types and intensities of distortions as training progresses, enabling the network to converge effectively and develop resilience to partial image theft. Comprehensive experiments demonstrate that our proposed scheme outperforms existing methods in scenarios of partial image theft, maintaining good visual quality and achieves high extraction accuracy across various image portions and backgrounds. Our results indicate that WOFA offers a solution for protecting digital images against unauthorized use in practical applications.

The main contributions of this paper are summarized as follows:

- We propose **WOFA** (Watermarking One for All), a robust watermarking scheme against partial image theft. It only needs embedding the watermark once into the image, and the watermark can be extracted from images that contains a portion of the original image. To the best of our knowledge, we are the first to introduce and conduct research on the scenario of partial image theft.
- We define the entire process of partial image theft and build a dataset accordingly. We design a progressive distortion layer which applies various types and intensities of distortions as training progresses, aiding the convergence of the network.
- Abundant experiments demonstrate that our proposed scheme outperforms existing methods in scenarios of partial image theft, providing a more reliable solution for protecting digital images against unauthorized use in practical use.

2. Related Works

Image Watermarking. Image watermarking has long been studied as an essential tool for copyright protection. Traditional watermarking schemes utilize well-crafted heuristic algorithms to embed watermarks in the spatial[1, 14] and frequency domains [2, 4, 9, 22], achieving robustness against common distortions such as JPEG compression[10], filtering[24], and geometric deformations[11]. With the advent of deep learning, numerous watermarking schemes leveraging neural networks have been developed. HiDDeN[30] was among the first to introduce an autoencoder-like architecture, jointly training an embedder, a distortion layer, and an extractor, which achieves both high robustness and good visual quality. Building on this architecture, methods such as LFM[25], PIMoG[6], and DeNoL[7] designed various distortion layers and training strategies to enhance robustness against screen-capturing, while StegaStamp[21], LIM[12], and WRAP[15] focused on improving robustness in the print-capture scenario. Additionally, MBRS[13] proposed a hybrid distortion layer to more effectively combat JPEG compression, DWSF[8] introduced a block-embedding scheme to ensure generalizability across different input resolutions and MuST[23] focuses on the challenge of image compositing and propose a method based on connected components to tackle it. Beyond the conventional embedder-distortion layer-extractor structure, UDH[27] pioneered a cover-independent watermarking approach inspired by the concept of universal adversarial examples, [28] enabled non-differentiable distortion layers by modifying the gradient propagation path, and CIN[19] integrated both invertible and non-invertible mechanisms within a watermarking framework.

3. Preliminaries

3.1. Definition of Partial Image Theft

In this section, we will define the entire process of partial image theft. In practical use, partial image theft occurs when malicious users steal only a portion of the original image, fuse it with a new background to form a new image. Therefore, we define the process of partial image theft as following steps and present it in Fig.2:

- **Partial Masking**. Partial masking refers to obtaining the

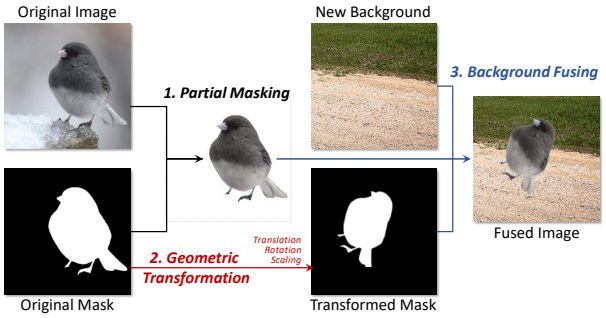

Figure 2. The entire process of partial image theft include 3 steps: partial masking, geometric transformation and background fusing.

stolen part $I_s \in \mathbb{R}^{3\times H\times W}$ from the original image $I_o \in \mathbb{R}^{3\times H\times W}$ based on a binary mask $m \in \{0,1\}^{1\times H\times W}$. In practical use, the mask is determined by the thief himself, depending on which part of the content the thief wants to steal.

- **Geometric Transformation**. In this step, the thief may perform geometric transformations on the stolen part to better fit the new background. In the scenario of partial image theft, translation is the most commonly used, followed by rotation and scaling.
- **Background Fusing**. Finally, the thief covers the stolen part on a new background $I_b \in \mathbb{R}^{3\times H\times W}$ to create a new image $I_f \in \mathbb{R}^{3\times H\times W}$. Without loss of generality, we assume in this paper that the size of the new background is the same as the original image.

3.2. Analysis to Partial Image Theft

In this section, we analyze the challenges that partial image theft poses to watermark extraction, aiming to design a more targeted solution. We summarize the impact of partial image theft into three points:

- **Content Loss**. The watermark information originally embedded in the entire image may become incomplete when only a portion is extracted.
- **Unknown Backgrounds**. The stolen part is fused with a background of unknown content that does not contain watermark, potentially interfering with the extraction.
- **Geometric Transformation**. Geometric transformation may change the position of the watermark information relative to the original image, which will also seriously interfere with the extraction.

Compared with the first two factors, geometric transformation has a greater impact on extraction. We design a simple experiment to reflect this effect. First, we obtain the original image, mask, and background image from the OPA[16] dataset, divide them into 5 groups according to the area ratio of the mask, and then perform the partial image theft process without any geometric transformation. As shown in Table.1, as the area of the mask (the stolen part) decreases, the extraction accuracy also decreases. This result shows the impact of content loss and unknown background on extraction. Then, we add a slight geometric transformation, only a 10% side length translation to the masks, and perform the partial image theft again. As shown in the table, for all area ratios, the extraction accuracy after translation has dropped significantly.

Table 1. Bit accuracy rate(BAR,%) of different mask area proportion. The ratio refers to the proportion of the mask's area relative to the original image.

Methods	Partial Masking(%)					Partial Masking(%) + 10% Translation				
	0-20	20-40	40-60	60-80	80-95	0-20	20-40	40-60	60-80	80-95
HiDDeN	52.89	62.90	71.35	78.53	84.84	52.81	61.05	67.06	71.01	73.05
MBRS	90.88	100	100	100	100	52.78	57.45	60.58	62.07	63.55
StegaStamp	60.73	79.07	89.92	96.17	98.85	52.01	55.85	59.25	62.06	64.41

We believe that the reason for this phenomenon may be that position information has a greater impact than content information. We further conduct another experiment to try to prove it. For each 128×128 image, we divide it into 16×16 blocks. First, we randomly set some of the blocks to all 0. Then, we change to not erasing any blocks, but shuffling some of the blocks, as shown in Fig.3. For the former, the position information is preserved but the content information is partially lost, while the latter is just the opposite. According to the proportion of blocks that did not change during the operation, we set 5 groups and test the extraction accuracy. As shown in Table.2 , for content loss, its performance decreases slowly as the loss amount increases, while for position information changes, its performance drops rapidly and significantly. This indicates that in the model design, an effective distortion layer must be designed to deal with the geometric transformation in partial image theft.

Table 2. Bit accuracy rate(BAR,%) under content information loss and position information loss. The ratio refers to the proportion of the number of the **unchanged** blocks in the process.

Methods	Block Erasing(%)					Block Shuffling(%)				
	0-20	20-40	40-60	60-80	80-95	0-20	20-40	40-60	60-80	80-95
HiDDeN	54.41	66.27	78.49	84.82	88.17	75.78	79.60	83.02	86.03	87.93
MBRS	94.04	100.00	100.00	100.00	100.00	70.53	94.83	99.90	100.00	100.00
PIMoG	64.69	80.49	90.32	96.57	99.39	55.95	68.27	80.81	91.88	98.08

Though factors above cause difficulties for extraction, there are still some methods can handle at least one of the above challenges. MBRS[13] introduced a diffusion block that disperses the watermark across the entire image, effectively addressing the issue of content loss. MuST[23] tried to tackle the challenge of image compositing and performs well against interference from unknown backgrounds. Ma et al.[18] adopts a transformer based model to combat geometric transformation. However, after careful study, we found none existing methods can address all three of the

Figure 3. Supplementary explanation for Section 3.2: After dividing the image into blocks, erasing and shuffling are applied to simulate content and position information loss, respectively.

above problems. Therefore, we try to propose a novel scheme for partial image theft.

4. Proposed Method

4.1. Overview

In order to provide a reliable solution for partial image theft, in this paper, we propose **WOFA** (Watermarking One for All). The structure of our model is illustrated in Fig.4. As depicted, our approach enhances the common-used embedder-distortion layer-extractor framework by introducing a multi-level architecture. During the embedding process, the watermark information is first mapped into a noise pattern by **encoder**, which is then embedded into the original image via the **embedder**. For extraction, the fused image created by malicious users is first processed by the **extractor** to retrieve the noise pattern, which is subsequently decoded into the watermark information by **decoder**. We will demonstrate in our experiments that this design facilitates easier convergence. For the model's core, the distortion layer, we employ a progressive strategy, applying various types and intensities of distortions at different stages of training to achieve optimal convergence.

As for the training strategy, in general, as depicted in Fig.4, we employ a two-stage training strategy. During the Stage I, we only focus on training the encoder and decoder. In the Stage II, we transit to joint training of the embedder and extractor within the full process, while keeping the weights of the encoder and decoder frozen. Each detail of training will be discussed in the following subsections.

4.2. Stage I: Training the Encoder & Decoder

As illustrated in Fig.4, in Stage I, we only focus on training the encoder and decoder. Encoder Enc maps the watermarking information $w \in \{0,1\}^{1\times l}$ to a grey-scale noise pattern $w_n^e \in \mathbb{R}^{1\times H\times W}$ of the same size as the original image, which can be formulated as:

$$w_n^e = Enc(w). \tag{1}$$

In the complete process, w_n^e will be then fed to the embedder to be embedded into the original image. However, during Stage I, we use a distortion layer to process w_n^e and then send it to the decoder Dec to output the predicted watermark $w_{pred} \in \{0,1\}^{1\times l}$, which can be formulated as:

$$w_n' = N_1(w_n^e), \tag{2}$$

$$w_{pred} = Dec(w_n'). \tag{3}$$

Here, N_1 refers to the distortion layer used exclusively in Stage I, which includes distortions such as partial masking and geometric transformations. The details of N_1 will be illustrated in following subsections.

The main purpose of Stage I is to build a mapping between the watermark information and the noise pattern as well as gaining the ability to resist content loss and geometric transformations. Considering that this mapping is relatively simple, we suppose that a simpler network architecture might offer better robustness and a reduced risk of overfitting. Therefore, we employ a light-weight design, utilizing two fully-connected (FC) layers for the encoder, and a decoder network consisting of four convolutional layers followed by two FC layers.

4.3. Stage II: Training the Embedder & Extractor

As shown in Fig.4, in Stage II, the complete process is utilized. The encoder first maps the watermark information $w \in \{0,1\}^{1\times l}$ to a grey-scale noise pattern $w_n^e \in \mathbb{R}^{1\times H\times W}$. Both the noise pattern w_n^e and the original image $I_o \in \mathbb{R}^{3\times H\times W}$ serve as the input of the embedder Emb to output the watermarked image $I_w \in \mathbb{R}^{3\times H\times W}$, which is formulated as:

$$I_w = Emb(w_n^e, I_o). \tag{4}$$

The watermarked image I_w is then processed by a distortion layer N_2 which incorporates not only the complete process of partial image theft but also geometric transformations and common perturbations such as JPEG compression:

$$I_f = N_2(I_w, I_b), \tag{5}$$

where $I_b \in \mathbb{R}^{3\times H\times W}$ refers to the background image and $I_f \in \mathbb{R}^{3\times H\times W}$ refers to the fused image constructed by covering a portion of I_w onto I_b. N_2 aims to enhance the model's robustness against unknown backgrounds and geometric transformations and its detail will be illustrated in the following subsection. Then, the extractor Ext will take the fused image I_f as input and output the predicted noise pattern w_n^p. Subsequently, the decoder will use w_n^p to predict the final watermark output w_{pred}, which can be formulated as:

$$w_n^p = Ext(I_f), \tag{6}$$

$$w_{pred} = Dec(w_n^p). \tag{7}$$

As illustrated in Fig.4, in the entire process of Stage II, the weight of encoder and decoder which are trained in Stage I is frozen, only embedder and extractor are being optimized. Considering that both the embedder and extractor

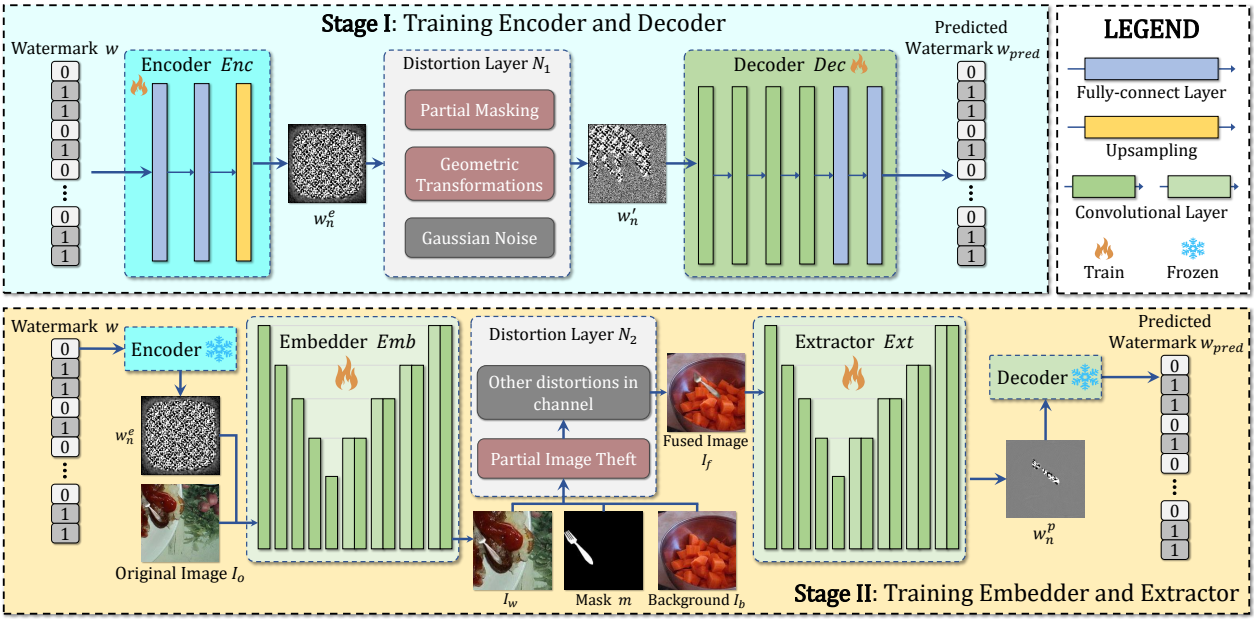

Figure 4. The flowchart of our proposed WOFA. WOFA adopts a multi-level structure, where a mapping between watermarking and noise pattern will be built. A two-stage training strategy is also utilized to aid convergence.

have the characteristic of having the same input and output shapes, we implement both using a U-Net structured[20] network with a depth of 5.

4.4. Distortion Layer

The distortion layer is usually crucial for achieving the model's robustness. In our approach, the distortion layer is primarily designed to address the three major challenges posed by partial image theft: content loss, unknown backgrounds, and the most important, geometric transformations. We apply different distortion layers N_1 and N_2 during Stage I and Stage II respectively and both of them are detailed in the following.

Distortion Layer I. Distortion layer N_1 is applied in Stage I and its main purpose is to gain robustness against content loss and geometric transformations. Therefore, N_1 consists of three different distortions as follows:

- **Partial Masking.** Partial masking is the first step of partial image theft, as shown in Fig.4. In the training of Stage I, we use partial masking to process the noise pattern w_n^e, which is equivalent to causing a content loss. To reduce the model's learning difficulty, we employ a gradual strategy. In the first 1% of training, we randomly mask squares covering at least 50% of the original pattern. During the next 10%, we transit to masking rectangles covering at least 25%. For the remaining training, we use masks from the dataset, with random, irregular shapes covering 1% to 95% of the image.
- **Geometric Transformations.** We have identified three key transformations in the scenario of partial image theft: translation, rotation and scaling, and incorporate them into our designed distortion layer. To ensure the effective area remains within the image boundaries after the transformation, we constrain the translation range to $\pm 50\%$ of the image's length or width and the scaling range to $\pm 25\%$. Additionally, to prevent excessive rotation from altering the semantics of the stolen portion, we limit the rotation range to $\pm 45°$. We also employ a progressive strategy for geometric transformations. The ranges for translation, rotation, and scaling gradually increase from 0 to their maximum values, reaching these limits at 50% of the training process.
- **Gaussian Noise.** In the complete process, the noise pattern to be decoded comes from the extractor rather than directly from the encoder in Stage I. Therefore, we introduce Gaussian noise to simulate any other possible distortions. The noise intensity (standard deviation) σ progressively increases from 0 to a maximum of 0.1 within the first 20% of training.

Distortion Layer II. Distortion layer N_2 is applied in Stage II to gain robustness against unknown backgrounds and geometric transformations. Additionally, this layer contributes to the robustness against other common distortions in digital channel. Consequently, N_2 consists of two different distortions as detailed below:

- **Partial Image Theft.** We define the complete process of partial image theft in Section 3.1 and incorporate its entire process as a distortion here. This process consists of

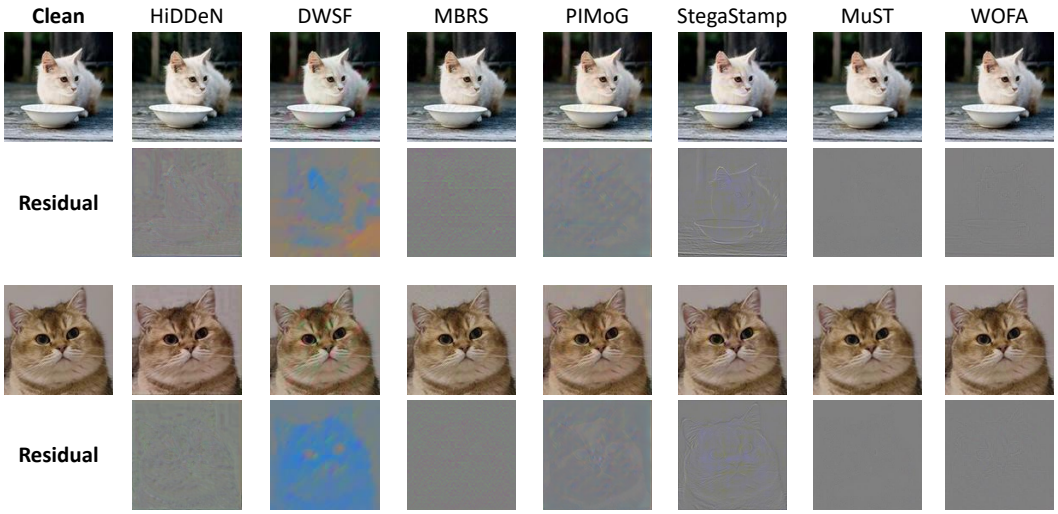

Figure 5. Visualization of watermarked images embedded by different methods. Our WOFA achieves satisfying visual performance overall.

three sub-processes: partial masking, geometric transformations, and background fusing. It is worth noting that unlike N_1, both partial masking and geometric transformations here do not apply gradual strategy but adopt a fixed intensity, that is the max intensity as illustrated in N_1.

- **Other Distortions in Digital Channel.** Although our scheme mainly focuses on the robustness against partial image theft, considering that the fused image may be transmitted through digital channels, we also incorporate JPEG compression (Quality Factor, QF = 95) which is a common channel distortion and resizing ($80\% - 125\%$) into this stage.

4.5. Training Strategy

For the training of our model, we adopt a two-stage strategy for better convergence. For Stage I, we only optimize the encoder and the decoder. We use only one Cross-Entropy loss function as constraint to minimize the distance between w and w_{pred}:

$$\mathcal{L}_1 = BCE(w, w_{pred}) \tag{8}$$

where $\mathcal{L}_1$ refers to the loss function in Stage I.

In Stage II, the parameters of both the encoder and the decoder are frozen. Only the embedder and the extractor need to be optimized. The optimization focuses on three goals: minimizing the difference between the original image I_o and watermarked images I_w, the noise pattern after geometric transformation w'_n and the predicted pattern w^p_n, and the original watermark w and predicted watermark w_{pred}. We utilize MSE as the loss function for the first two, while the third goal follows the loss function in Stage I, which can be formulated as:

$$\mathcal{L}_{image} = MSE(I_o, I_w), \tag{9}$$

$$\mathcal{L}_{pattern} = MSE(w'_n, w^p_n), \tag{10}$$

$$\mathcal{L}_2 = \mathcal{L}_{image} + \mathcal{L}_{pattern} + \mathcal{L}_1, \tag{11}$$

where $\mathcal{L}_2$ refers to the loss function in Stage II.

5. Experiments

5.1. Experimental Setting

5.1.1. Datasets

For the experiments, we select several different datasets. First, we sample 61,990 images from OPA[16], which is a dataset about object placement and fusing. We crop blocks with size 200×200 and then split them into a training set and a test set with a 4:1 ratio. All the baseline models and our model are trained on this training set and most of the experiments in this paper are conducted on this test set except for the generalizability test. Secondly, we also test our model on SOIM[23] and matteImageNet to demonstrate the generalization of our method. MatteImageNet is the dataset we constructed based on the process of partial image theft. We use MatteAnything[26] to obtain 977 images and their corresponding masks from 19 selected classes in ImageNet[5].

5.1.2. Implementation Details.

Our method is implemented by PyTorch [3] and trained on a single NVIDIA RTX 3090ti. We use the Adam[17] optimizer with a learning rate of 5e-5. The length l of watermark w is set as 30. All the methods we used for comparison [6, 8, 13, 21, 23, 30] in experiments are re-trained using their open-source code and the configurations in their papers.

Table 3. Visual quantification in different methods

Methods	Image size	Capacity	Bits per pixel	PSNR↑	SSIM↑	LPIPS↓
HiDDeN	128×128	30 bits	1.83×10^{-3}	30.37	0.9247	0.0493
DWSF	128×128	30 bits	1.83×10^{-3}	38.89	<u>0.9879</u>	**0.0102**
MBRS	128×128	30 bits	1.83×10^{-3}	32.82	0.8827	0.0876
PIMoG	128×128	30 bits	1.83×10^{-3}	33.66	0.9625	0.0189
StegaStamp	400×400	100 bits	6.25×10^{-4}	27.69	0.9088	0.1123
MuST	256×256	30 bits	4.58×10^{-4}	**43.31**	0.9796	0.0704
WOFA	200×200	30 bits	7.50×10^{-4}	<u>39.76</u>	**0.9945**	<u>0.0188</u>

5.2. Visual Performance

To evaluate the visual quality of our method, we compare the visual quality of watermarked images obtained through various methods. As shown in Table.3, we compare our method with several SOTA watermarking schemes on three metrics, namely Peak Signal-to-Noise Ratio (PSNR), Structure Similarity Index Measure (SSIM) and Learned Perceptual Image Patch Similarity (LPIPS) [29]. As shown in Table.3, our method achieves commendable PSNR (nearly 40 dB) and SSIM, and has an advantage in LPIPS which better reflects the visual perception quality. We also present a more intuitive result in Fig.5 which displays watermarked images generated by different methods. The visual comparison further confirms that our model's image quality is highly satisfactory.

5.3. Robustness Performance in the Scenario of Partial Image Theft

In this section, we evaluate the performance of our method against partial image theft by comparing it with several state-of-the-art (SOTA) pre-trained models. First of all, we test the extraction performance of the models under partial masking with different area proportion ratios to observe the resistance to content loss and unknown background. As shown in Table.4, our model achieve a good performance which demonstrates that WOFA is able to resist both content loss and unknown backgrounds.

We then evaluate the performance against all three main challenges. Table 5 presents the bit accuracy results under different settings. First, we present the performance under the "partial" setting, which involves only partial masking of area proportion ranged from 1% to 95% and without geometric transformations. We then combine partial masking with various geometric transformations, including combinations of transformations to better simulate real-world scenarios. As shown in the table, our WOFA method consistently outperforms other models, achieving over 90% bit accuracy across most settings. This demonstrates WOFA's superior robustness against geometric transformations and effectiveness in handling partial image theft.

Additionally, we evaluate the robustness of our model against common distortions encountered in digital channels, and the results are presented in Appendix since the space is limited. Although these distortions are not the primary focus of our model, WOFA still performs well across a range of different distortions, maintaining a high level of extraction accuracy. Overall, our model proves to be highly effective in scenarios involving partial image theft and offers a more reliable solution for digital watermarking in practical applications.

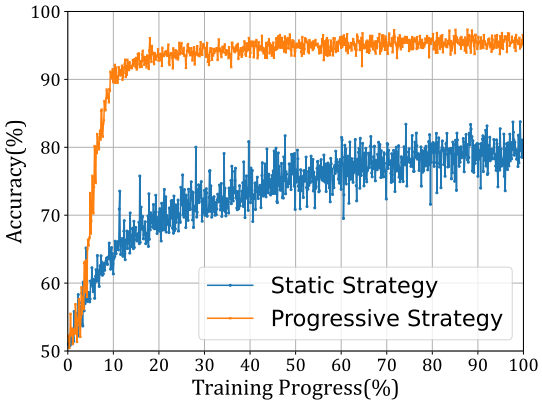

Figure 6. The training process under two different strategies. The progressive distortion layer achieves faster and better convergence.

5.4. Generalizability to Different Datasets

The previous experiments were all conducted on the OPA dataset. Considering that our model is also trained on the training set constructed by OPA, in order to verify the generalization ability of our model, in this section, we evaluate the model's generalization performance on various datasets. In addition to the OPA dataset, we conduct experiments on the SOIM dataset and our self-constructed matteImageNet dataset. SOIM is a dataset for image compositing and most of its images contain only a single object and a simple background. MatteImageNet is the dataset we built on the basis of ImageNet, which offers more diverse foreground and background. The results are shown in Table 6. The results demonstrate that our model maintains strong performance across these diverse datasets, highlighting its ability to generalize effectively to different types of data. This robustness across multiple datasets emphasizes the versatility and adaptability of our approach, making it a reliable choice for various real-world applications.

5.5. Ablation Study

Progressive Distortion Layer. During the training process, we employ a progressive distortion layer that gradually increases the intensity of distortions as training advances. To demonstrate the benefits of this strategy on convergence, we performed an ablation study, with the results presented in Fig.6. As shown in the figure, the progressive strategy enables the network to converge more quickly and ultimately reach an optimal point with improved performance, which

Table 4. Bit accuracy rate(BAR,%) under partial masking of different area proportion.

Methods	Clean	Partial Masking(%)									
		0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
HiDDeN	90.51	51.75	55.79	60.54	64.89	69.27	73.45	77.19	78.96	82.73	86.41
DWSF	99.99	49.97	49.85	50.17	51.09	53.61	56.01	61.94	69.60	75.17	77.02
MBRS	100	85.96	**99.80**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
PIMoG	99.92	55.54	66.11	73.94	81.19	88.52	92.08	95.78	98.43	99.46	99.86
StegaStamp	98.93	55.18	66.28	75.74	82.40	88.12	91.72	95.23	97.11	98.54	99.16
MuST	99.59	50.03	50.01	49.78	50.62	51.00	50.24	50.80	50.67	50.97	51.65
WOFA	98.39	**89.33**	94.90	96.42	97.61	98.23	98.55	98.61	98.66	98.62	98.67

Table 5. Bit accuracy rate(BAR,%) under different distortions in Partial Image Theft. (TS refers to translation for 50% side length, RT refers to rotation for 45°, SC refers to scaling for $\pm 25\%$. Mixed 3 refers to the mixed transformations of this three.

Methods	Clean	Partial	Partial + Translation(%)			Partial + Rotation(°)			Partial + Scaling(%)			Partial + Mixed 2			Partial + Mixed 3
			10	25	50	15	30	45	±10	±20	±25	TS + RT	TS + SC	RT + SC	
HiDDeN	90.51	62.47	59.67	58.93	56.99	59.14	56.89	54.20	59.73	59.29	58.91	53.07	55.98	53.97	52.59
DWSF	99.99	55.84	52.97	49.87	49.92	50.21	49.74	49.60	53.30	51.78	51.08	49.84	49.66	49.71	49.91
MBRS	100	94.74	56.19	59.13	52.15	50.43	50.17	49.56	50.27	50.14	49.91	50.73	49.88	49.58	49.24
PIMoG	99.92	74.91	57.72	47.59	51.30	49.31	49.83	50.02	62.58	50.16	49.52	49.47	50.43	50.03	49.96
StegaStamp	98.93	74.48	55.85	49.47	49.83	56.92	50.68	50.04	70.19	61.54	58.36	49.86	49.78	50.07	49.97
MuST	99.59	52.77	50.00	49.98	49.74	49.79	49.73	49.82	49.98	49.99	50.00	49.93	50.15	50.04	50.13
WOFA	98.39	**95.92**	**91.97**	**93.25**	**87.93**	**95.26**	**94.24**	**90.63**	**95.72**	**95.50**	**95.02**	**87.20**	**91.34**	**90.74**	**87.06**

Table 6. Bit accuracy rate(BAR,%) of WOFA against Partial Image Theft tested on different datasets

Dataset	Clean	Partial	Partial + Translation	Partial + Rotation	Partial + Scaling	Partial + Mixed 3
OPA	98.39	95.92	87.93	90.63	95.02	87.06
matteImageNet	94.44	89.78	83.98	83.01	92.91	85.05
SOIM	96.94	90.56	89.99	86.37	94.08	86.52

demonstrates the effectiveness of our design.

Multi-stage Training. In our training process, we employ a two-stage training strategy rather than an end-to-end approach. We perform an ablation study to demonstrate why this multi-stage training strategy is better. Since in our experiments, the end-to-end training scheme failed to converge, making it impossible to directly compare convergence speed and final performance. Instead, we present intermediate results from the training process in Fig.7. As shown, the end-to-end approach will collapse early in training (around 10% of the total training time), leading to nearly blank outputs and extraction accuracy that stagnates around 50%. In contrast, our multi-stage strategy steadily approaches the ground truth and ultimately achieves better convergence.

6. Conclusion

In this paper, we introduce WOFA (Watermarking One for All), a robust watermarking scheme designed to combat partial image theft, a scenario where only a portion of an image is stolen or misused. To address the three main challenges brought by the scenario: content loss, unknown backgrounds and geometric transformations, we design a multi-level structured network and meticulously design a progressive distortion layer applying various types and intensities of distortions as training progresses, enabling the network to converge effectively and develop resilience to partial image theft. Extensive experiments demonstrate that WOFA outperforms existing watermarking schemes by effectively addressing the practical challenges posed by partial image theft, offering a reliable solution for digital image protection in today's Internet-driven world.

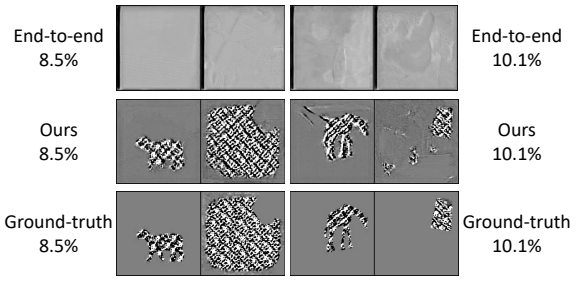

Figure 7. The prediction of noise pattern from different schemes. The end-to-end scheme usually collapse and fail to convergence at very early of the training, leading to a nearly blank output.

Acknowledgment. This work was supported by the Natural Key R&D Program of China under Grant 2023YFF0905000.

References

[1] Gordon W Braudaway. Protecting publicly-available images with an invisible image watermark. In *Proceedings of international conference on image processing*, pages 524–527. IEEE, 1997. 2

[2] Wai C Chu. Dct-based image watermarking using subsampling. *IEEE transactions on multimedia*, 5(1):34–38, 2003. 2

[3] Ronan Collobert, Koray Kavukcuoglu, and Clément Farabet. Torch7: A matlab-like environment for machine learning. In *BigLearn, NIPS workshop*, 2011. 6

[4] Ingemar J Cox, Joe Kilian, F Thomson Leighton, and Talal Shamoon. Secure spread spectrum watermarking for multimedia. *IEEE transactions on image processing*, 6(12):1673–1687, 1997. 2

[5] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 6

[6] Han Fang, Zhaoyang Jia, Zehua Ma, Ee-Chien Chang, and Weiming Zhang. Pimog: An effective screen-shooting noise-layer simulation for deep-learning-based watermarking network. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 2267–2275, 2022. 2, 6

[7] Han Fang, Kejiang Chen, Yupeng Qiu, Jiayang Liu, Ke Xu, Chengfang Fang, Weiming Zhang, and Ee-Chien Chang. Denol: A few-shot-sample-based decoupling noise layer for cross-channel watermarking robustness. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 7345–7353, 2023. 2

[8] Hengchang Guo, Qilong Zhang, Junwei Luo, Feng Guo, Wenbin Zhang, Xiaodong Su, and Minglei Li. Practical deep dispersed watermarking with synchronization and fusion. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 7922–7932, 2023. 2, 6

[9] Mingze He, Hongxia Wang, Fei Zhang, and Yuyuan Xiang. Exploring accurate invariants on polar harmonic fourier moments in polar coordinates for robust image watermarking. *IEEE Transactions on Multimedia*, 2023. 2

[10] Chiou-Ting Hsu and Ja-Ling Wu. Hidden digital watermarks in images. *IEEE Transactions on image processing*, 8(1):58–68, 1999. 2

[11] Runwen Hu and Shijun Xiang. Cover-lossless robust image watermarking against geometric deformations. *IEEE Transactions on Image Processing*, 30:318–331, 2020. 2

[12] Jun Jia, Zhongpai Gao, Dandan Zhu, Xiongkuo Min, Guangtao Zhai, and Xiaokang Yang. Learning invisible markers for hidden codes in offline-to-online photography. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2273–2282, 2022. 2

[13] Zhaoyang Jia, Han Fang, and Weiming Zhang. Mbrs: Enhancing robustness of dnn-based watermarking by mini-batch of real and simulated jpeg compression. In *Proceedings of the 29th ACM international conference on multimedia*, pages 41–49, 2021. 2, 3, 6, 1

[14] Irene G Karybali and Kostas Berberidis. Efficient spatial image watermarking via new perceptual masking and blind detection schemes. *IEEE Transactions on Information Forensics and security*, 1(2):256–274, 2006. 2

[15] Gaozhi Liu, Yichao Si, Zhenxing Qian, Xinpeng Zhang, Sheng Li, and Wanli Peng. Wrap: Watermarking approach robust against film-coating upon printed photographs. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 7274–7282, 2023. 2

[16] Liu Liu, Zhenchen Liu, Bo Zhang, Jiangtong Li, Li Niu, Qingyang Liu, and Liqing Zhang. Opa: object placement assessment dataset. *arXiv preprint arXiv:2107.01889*, 2021. 3, 6

[17] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. 6

[18] Linfeng Ma, Han Fang, Tianyi Wei, Zijin Yang, Zehua Ma, Weiming Zhang, and Nenghai Yu. A geometric distortion immunized deep watermarking framework with robustness generalizability. 3

[19] Rui Ma, Mengxi Guo, Yi Hou, Fan Yang, Yuan Li, Huizhu Jia, and Xiaodong Xie. Towards blind watermarking: Combining invertible and non-invertible mechanisms. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 1532–1542, 2022. 2

[20] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015. 5

[21] Matthew Tancik, Ben Mildenhall, and Ren Ng. Stegastamp: Invisible hyperlinks in physical photographs. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2117–2126, 2020. 2, 6, 1

[22] Huawei Tian, Yao Zhao, Rongrong Ni, Lunming Qin, and Xuelong Li. Ldft-based watermarking resilient to local desynchronization attacks. *IEEE Transactions on cybernetics*, 43(6):2190–2201, 2013. 2

[23] Guanjie Wang, Zehua Ma, Chang Liu, Xi Yang, Han Fang, Weiming Zhang, and Nenghai Yu. Must: Robust image watermarking for multi-source tracing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 5364–5371, 2024. 2, 3, 6, 1

[24] Xiang Wang, Xiaolong Li, and Qingqi Pei. Independent embedding domain based two-stage robust reversible watermarking. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(8):2406–2417, 2019. 2

[25] Eric Wengrowski and Kristin Dana. Light field messaging with deep photographic steganography. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1515–1524, 2019. 2

[26] Jingfeng Yao, Xinggang Wang, Lang Ye, and Wenyu Liu. Matte anything: Interactive natural image matting with segment anything model. *Image and Vision Computing*, 147: 105067, 2024. 6

[27] Chaoning Zhang, Philipp Benz, Adil Karjauv, Geng Sun, and In So Kweon. Udh: Universal deep hiding for steganography, watermarking, and light field messaging. *Advances in Neural Information Processing Systems*, 33:10223–10234, 2020. 2

[28] Chaoning Zhang, Adil Karjauv, Philipp Benz, and In So Kweon. Towards robust deep hiding under non-differentiable distortions for practical blind watermarking. In *Proceedings of the 29th ACM international conference on multimedia*, pages 5158–5166, 2021. 2

[29] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 7

[30] Jiren Zhu, Russell Kaplan, Justin Johnson, and Li Fei-Fei. Hidden: Hiding data with deep networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 657–672, 2018. 2, 6