

Article

An Improved U-Net for Watermark Removal

Lijun Fu ¹, Bei Shi ², Ling Sun ³, Jiawen Zeng ², Deyun Chen ^{1,*}, Hongwei Zhao ⁴ and Chunwei Tian ^{2,5} 

¹ School of Computer Science and Technology, Harbin University of Computer Science and Technology, Harbin 150080, China

² School of Software, Northwestern Polytechnical University, Xi'an 710129, China

³ School of Information and Electromechanical Engineering, Heilongjiang University of Industry and Business, Harbin 150025, China

⁴ Shandong Baimeng Information Technology Co., Ltd., Weihai 264200, China

⁵ Research & Development Institute, Northwestern Polytechnical University, Shenzhen 518000, China

* Correspondence: tg1950@hrbust.edu.cn

Abstract: Convolutional neural networks (CNNs) with different layers have performed with excellent results in watermark removal. However, how to extract robust and effective features via CNNs of black box in watermark removal is very important. In this paper, we propose an improved watermark removal U-net (IWRU-net). Taking the robustness of obtained information into account, a serial architecture is designed to facilitate useful information for guaranteeing performance in watermark removal. Taking the problem of long-term dependency into account, U-nets based simple components are integrated into the serial architecture to extract more salient hierarchical information for addressing watermark removal problems. To increase the adaptability of IWRU-net to the real world, we use randomly distributed blind watermarks to implement a blind watermark removal model. The experiment results illustrate that the proposed method is superior to other popular watermark removal methods in terms of quantitative and qualitative evaluations.

Keywords: serial architecture; U-net; blind watermark removal



Citation: Fu, L.; Shi, B.; Sun, L.; Zeng, J.; Chen, D.; Zhao, H.; Tian, C. An Improved U-Net for Watermark Removal. *Electronics* **2022**, *11*, 3760. <https://doi.org/10.3390/electronics11223760>

Academic Editor: Stefanos Kollias

Received: 16 October 2022

Accepted: 10 November 2022

Published: 16 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

To protect the copyright of files, added watermarks became a popular way to increase security of protected files [1]. To test the quality of added watermarks, watermark removal is an effective tool [2]. Integrating prior occurrences and the likelihood of cross-channel correlation can repair image information of the damaged channel in watermark removal [3]. Just-noticeable-distortion is used to estimate the energy to remove the watermark [4]. Taking into questions of scanned and back-lit pages in archaic documents account, a known lexicon of fragments is exploited to find watermarks and remove them for overcoming the effects of damaged files [5]. To improve the effect of removing watermarks, the discrete cosine transform domain and a key-based matrix are fused to remove visible watermarking [6]. Alternatively, authors use entropy and edge entropy as human visual system (HVS) characteristics to quickly extract watermark features and remove them [7]. To improve the generalization ability of the watermark removal algorithm, using wavelet transform can extract watermark features, the obtained features are used to revise singular values in order to test the robustness of the obtained watermark [8]. Ansari et al. adjusted block differences between HL and LH to automatically select the wavelet coefficient for improving watermark quality [9]. In addition, Fourier transform is effective for watermark removal. For instance, using the Fourier transform domain to remove watermarks of R, G and B in the watermark images is a good tool for blind color image watermark removal [10]. Taking robustness and imperceptibility into account, scholars ensure a signal-to-noise ratio that enables the host image to maximize the embedding strength for making a tradeoff between robustness and imperceptibility in the image watermarking removal [11]. Although these methods have performed well in watermark removal, they still suffer from the following

drawbacks: (1) They use complex optimization methods to improve visual effect of watermark removal. (2) Manually choosing parameters is a good tool to improve the performance of watermark removal. To overcome these disadvantages, we use deep learning techniques, especially convolutional neural networks to deal with watermark removal.

Chen et al. proposed deep neural networks for watermark removal [12]. To improve the quality of watermark removal, Sai et al. used lower dimensional projections in the intermediate layers of a deep CNN to express the image content in watermark removal [13]. Haribabu et al. utilized an auto-encoder to deal with watermark images with two independent images for watermark removal [14]. To improve the robustness of watermarks, Chen et al. used an adaption of elastic weight consolidation and unlabeled data augmentation to better represent watermarks for watermark removal [15]. Using roughly localized and separate watermarks is a good tool for an image watermark [16]. Using a CNN, wavelet transform and residual regularization loss function, rather than down- and up-sampling operations, can improve the visual quality of watermark images [17].

Although these CNNs have obtained comparative results in watermark removal, the key question is how to extract effective features of CNNs with black box to better represent watermarks for more complex watermark removal. In this paper, we propose an improved watermark removal U-net (IWRU-net). To obtain more robust information, a serial architecture is presented to extract useful information to pursue better performance for watermark removal. To address a long-term dependency problem, U-net's base simple components are fused into the designed serial architecture to extract more salient hierarchical information for dealing with the watermark removal problem. Taking into the adaptability of IWRU-net in the real world account, we use randomly distributed blind watermarks to conduct a blind watermark removal model. The experiment's results show that the proposed method is effective in quantitative and qualitative evaluations.

This paper makes the following contributions.

1. A serial architecture is used to facilitate more useful information for improving the performance of watermark removal.
2. U-nets are gathered into a serial architecture to extract more salient hierarchical information to address the long-term dependency on deep CNNs for watermark removal.
3. To improve the adaptability of IWRU-net on mobile devices in the real world, randomly distributed watermarks with different types are used to train a blind watermark removal model.

The remaining parts of this paper have the following organizations. Section 2 represents related work on the proposed method. Section 3 lists the proposed method. Section 4 presents experiments. Section 5 offers conclusions.

2. Related Work

2.1. Deep CNNs for Watermark Removal

Due to their strong learning ability, CNNs are often exploited for image applications [18,19]. For example, CNNs are used to extract robust features for better representing watermarks for watermark removal [20]. To address different watermarks, a detector based on the CNN is used to detect the locations of watermarks and to remove watermarks [20]. To deal with blind watermark removal, Lee et al. used a pre-processing network, a watermark embedding network and a watermark extraction network to enhance the host image with super-resolution and the watermark invisibility for blind watermark removal [21]. To address artifacts and blurriness caused by opaque watermarks, dual convolutions are used for watermark removal [22]. That is, the first network is exploited to remove the watermark. The second network is utilized to optimize the second network for further filter watermarks. Due to problem of conventional perceptual hashing, perceptual hash is embedded into a CNN to obtain a weight way for verifying watermarking [23]. To address watermarks in remote operations, a novel zero-bit watermarking algorithm with adversarial model examples was used to extract the watermark and remove it [24]. In a tradeoff between robustness and transparency, a combination

of water wave optimization, chaotic fruit fly optimization algorithm and CNN was developed for verifying watermarks [25]. To discriminate between attack watermarks and corrected watermarks, a generative adversarial network (GAN) was used for watermark removal [26]; that is, a generative network based on U-net architecture was used to extract high- and low-level features for generating images. Also, a discriminative network was used to distinguish between the truth of generated and original images, which can also be used to remove the watermark. To reduce the time of watermark removal, a lightweight CNN was designed [27]. To improve watermark removal, a discrete cosine transformation and Harris hawks method were fused into a CNN to filter watermarks [28]. To improve the ability of watermark removal, a progressive pre-processing operation was gathered into a residual dense network to extract more low-frequency features and enhance the attack ability of the designed network for image watermark removal [29]. Motivated by that, we designed a novel CNN for image watermark removal.

2.2. Cascaded Architectures for Image Applications

To extract richer features, a cascaded architecture was designed to improve the performance in image applications [30,31]. For instance, Qin et al. trained a cascaded network composed simultaneously of a region proposal network and a fast R-CNN to improve the accuracy rate of facial detection [30]. To address undersampled data, a cascaded CNN was presented with MR images for overcoming undersampled data [32]. To overcome the effect of noise and artifacts, two identical networks were cascaded to enhance the classification results of medical images [33]. That is, the first network was used to remove noise; the second network was exploited to classify medical images. Taking into effect of different factors for image dehazing account, Li et al. used two sub-networks to address medium transmission and global atmospheric light to obtain more realistic effects, closer to the real world, to improve the practicality of the proposed method [34]. Alternatively, Yan et al. combined multi-frame geometry and jointed training to gather low- and high-frequency information to enhance the image quality of consumer depth cameras [35]. To improve the performance of image registration, each cascaded network can further deal with warped images to change the image quality [36]. To fully deal with the specific attributes of HSIs, a cascaded architecture based on two recurrent neural networks was used for hyper-spectral image classification [37]. Specifically, the first RNN was applied to remove redundant information from spectral bands. The second RNN can extract more extra information obtained from nonadjacent spectral bands. Two networks can improve the discriminative ability of the obtained classifier. Besides, a cascaded network can use a hierarchical architecture to extract more useful features to enhance the image quality [38]. Tian et al. designed a heterogeneous architecture and a stacked convolutional layer to mine richer low- and high-frequency features for addressing the unstable problem of a SR model [39]. To overcome the challenge of the low spatial resolution of hyper-spectral images, a network was implemented by cascading two sub-networks [40]. That is, the first network was used to obtain high resolution multispectral panchromatic images; the second network was exploited to predict abundance maps. Two networks can better deal with hyperspectral image resolution. The cascade network architecture was extended for facial expression recognition [41]. Combining group convolutional networks and stacked CNNs can be used to enhance the relationships of different channels for image super-resolution [42,43]. Following the above studies, we can see that cascading networks are useful for image applications. Inspired by that, we designed a cascading network architecture for image watermark removal.

3. The Proposed Method

3.1. Network Architecture

To extract robust and effective features for watermark removal, we propose an improved watermark removal U-net with 42 layers as well as an IWRU-net, as reported in Figure 1. To improve the robustness of obtained features, a serial architecture is im-

plemented by cascading two sub-networks in order to obtain effective information for improving the performance of watermark removal. To address long-term dependency problem, U-nets base simple components are fused into the designed serial architecture to extract more salient hierarchical information for dealing watermark removal problem. To increase the adaptability of the IWRU-net to the real world, we use randomly distributed blind watermarks to implement a blind watermark removal model. To roughly express the above, we conduct the following equation.

$$\begin{aligned} I_c &= IWRU_{net}(I_w) \\ &= UnetBlock(UnetBlock(I_w)) \end{aligned} \quad (1)$$

where I_w denotes a watermark image and $IWRU_{net}$ is a function of IWRU-net. I_c represents a clean image. $UnetBlock$ expresses the function of the Unet Block. Also, $IWRU_{net}$ is trained by the loss function illustrated in Section 3.2. Finally, each Unet Block is shown in Section 3.3.

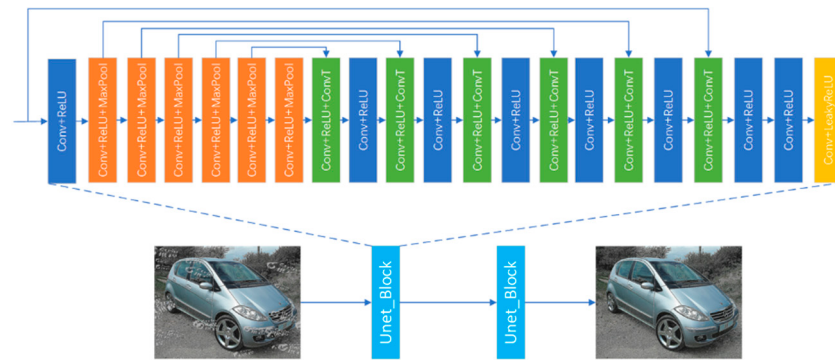


Figure 1. Network architecture of IWRU-net.

3.2. Loss Function

To improve the training efficiency, least absolute deviation (LAD) [44,45] is used to train the IWRU-net model for image watermark removal. That is, we use a watermark image and a clean image to act in an IWRU-net according to Equation (2) to train a watermark removal model.

$$D(\theta) = \frac{\sum_{j=1}^t |I_w^j - IWRU_{net}(I_w^j)|}{t} \quad (2)$$

where I_w^j is the j th watermark image, t expresses the total number of watermark images, D denotes a loss function for training a IWRU-net watermark removal model and θ is used to represent the parameters. Note that the parameters are optimized by Adam [46] when the IWRU-net is trained.

3.3. Each Unet Block

A serial architecture composed of two sub-networks is key to improving the robustness of the obtained features in the IWRU-net. Also, each 21-layer U-net filling in each sub-network is used to obtain more salient hierarchical information for watermark removal. That is, the 1st, 9th, 11th, 13th, 15th, 17th, 19th and 20th layers are composed of Conv+ReLU; the 2nd–7th layers include Conv+ReLU+MaxPool; the 8th, 10th, 12th, 14th, 16th and 18th layers contain Conv+ReLU+ConvT. Also, the last layer includes Conv+LeakyReLU. The mentioned Conv+ReLU is a combination of a convolutional layer and an activation function of ReLU [47]. The mentioned Conv+ReLU+MaxPool is a combination of a convolutional layer, an activation function of ReLU and a pooling function of max pooling [48]. Conv+ReLU+ConvT is a combination of a convolutional layer, ReLU and deconvolution

(ConvTranspose2d). In addition, a convolutional layer is used to obtain linear features. ReLU is exploited to map linear features into non-linear features. MaxPool is utilized to reduce the dimension of data to improve the efficiency of training a IWRU-net model. ConvTranspose2d is exploited to obtain our predicted results. LeakyReLU is exploited to map linear features onto non-linear features [49]. To enhance the memory ability of IWRU-net, we use concatenation operations to integrate shallow features to transmit deep layers. That is, features of the input and the 18th layer are fused by a concatenation operation as an input of the 19th layer. Features of the 2nd layer and 16th layer are merged by a concatenation operation as an input of the 16th layer. Features of the 3rd layer and 14th layer are gathered by a concatenation operation as an input of the 15th layer. Features of the 4th layer and 12th layer are fused by a concatenation operation as an input of the 13th layer. Features of the 5th layer and 10th layer are gathered by a concatenation operation as an input of the 11th layer. Features of the 6th layer and 8th layer are gathered by a concatenation operation as an input of the 9th layer. Also, each convolutional kernel size is 3×3 . Input and output channel number of each layer are shown as follows. The 1st layer has 3 input channels and 48 output channels. The 2nd–8th layers have 48 input and output channels, respectively. The 9th, 10th, 12th, 14th, 16th and 18th layers have 96 input and output channels. The 13th, 15th and 17th have 144 input channels and 96 output channels, respectively.

The 19th layer has 99 input channel and 64 output channels. The 20th layer has 64 input channels and 32 output channels. The 21st layer has 32 input channels and 3 output channels.

To allow readers to understand illustrations the above visually, we conducted the following equations.

$$\begin{aligned} O_{F_UnetBlock}^i &= UnetBlock(I_w) \\ &= CLR(CR(CR(C_0(CRC(CR(C_0(CRC(CR(C_0(CRC(CR(C_0(CRC(O_t), O_4))), O_3))), O_2))), I_w)))) \end{aligned} \quad (3)$$

$$O_t = CR(C_0(CRC(CR(C_0(CRC(6MCR(CR(I_w))), O_6))), O_5)) \quad (4)$$

$$O_6 = 5MCR(CR(I_w)) \quad (5)$$

$$O_5 = 4MCR(CR(I_w)) \quad (6)$$

$$O_4 = 3MCR(CR(I_w)) \quad (7)$$

$$O_3 = 2MCR(CR(I_w)) \quad (8)$$

$$O_2 = MCR(CR(I_w)) \quad (9)$$

where CR denotes a combination of a convolution and ReLU; $nMCR$ is n stacked MCR, where n varies from 1 to 6; CRC expresses a combination of a convolution, ReLU and Max pooling; CLR represents a combination of a convolution and LeakyReLU; O_j stands for output of the j th layer and $j = 2, 3, 4, 5, 6$; O_t is a temporary output; $O_{F_UnetBlock}^i$ is the i th output of the Unet Block ($i = 1, 2$).

4. Experiments

4.1. Datasets

Training dataset. Following [50–52], we chose public large-scale visible watermarks (LVW) [20] for our training dataset to train our IWRU-net. The training dataset is composed of 60,000 watermarked images with 80 watermarks. Each watermark is embedded into 750 images. Also, we chose 3000 images without watermarks from the LVW. To enlarge the categories of training samples, we rotated seven conducted watermarks from -30° to 30° , scaled them from 70% to 100% and adjusted them to a transparency of 50% to 80% then randomly added them to the mentioned 3000 images to achieve a watermark coverage of 10% for constructing watermark images, where watermark coverage is the ratio of watermarking pixels to all the pixels in an entire image.

Test datasets. In order to fairly test the performance of our IWRU-net for watermark removal, we randomly selected 200 images from the LVW and colored large-scale watermark dataset (CLWD) [16] as test datasets. Specifically, 100 images were chosen from the LVW and the rest were chosen from the CLWD. Watermark images were rotated from -30° to 30° , underwent random scaling of 70% to 100% and random transparency adjustments of 50% to 80%. Additionally, a watermark was added into an image as a test image.

4.2. Experimental Settings

Our experiments were conducted on a PC. The PC has two CPUs of Intel(R) Xeon(R) Silver 4210 CPU@2.20GHz with RAM of 128 G and a Nvidia GeForce GTX 3090 GPU, where CUDN of 11.1 and CUDNN of 7.4.1 are used to accelerate the GPU. Our codes are run by PyTorch of 1.10.2 and Python of 3.8.12 on Ubuntu of 20.04.2. The initial learning rate is set to 1×10^{-4} . The number of epochs is 100. The learning rate has 0.5 times reduction each 200,000 iterations. All the training images and test images were scaled to 512×512 as inputs of the IWRU-net. Outputs of the IWRU-net need to be scaled the same as the original images. Other parameters are the same as in [53].

4.3. Experimental Analysis

Due to their hierarchical architecture, deep CNNs have obtained stronger learning abilities for image application, where hierarchical features have obtained features from different layers [54]. However, how to ensure obtained robust features is very important. Following Section 2.2, we can see that cascaded architectures are suitable to mine for more accurate features for image applications. Inspired by that, we designed an improved watermark removal U-net (IWRU-net) based on a cascaded architecture. That is, a serial architecture is used to facilitate useful information for guaranteeing performance in watermark removal. In this paper, we used two blocks (Unet_Blocks) in a serial way to form the serial architecture for image watermark removal in Figure 1. To address long-term dependency problems, we chose U-net as a simple component of each block (Unet_Block) to extract more salient hierarchical information to address watermark removal problems. To verify the effectiveness of the serial architecture, we used an IWRU-net and a single U-net on 100 images with seven watermarks from the LVW dataset to conduct experiments, where the settings of the chosen watermark images are the same as in Section 4.2. That is, the IWRU-net obtained higher peak signal-to-noise ratio (PSNR) [55] than that of a single U-net for image watermark removal as shown in Table 1, which shows the effectiveness of serial architecture in image watermark removal.

To increase the diversity of the obtained features, we used six up-sampling and down-sampling operations in each U-net to extract richer features. We used IWRU-net and IWRU-net without up- and down-sampling operations to test the superiority of up- and down-sampling operations in the IWRU-net for image watermark removal as reported in Table 1. To test the effectiveness of six up- and down-sampling operations in each U-net for image watermark, we chose the IWRU-net and the IWRU-net with four up- and down-sampling operations in each U-net to conduct comparative experiments in Table 1. That shows that the proposed IWRU-net exceeds IWRU-net with four up- and down-sampling operations in each U-net in terms of PSNR, which verifies the good performance of the six up- and down-sampling operations. Also, six up- and down-sampling operations can enhance the expressive ability of the designed IWRU-net. To discuss the effect of residual operations on serial architecture for image watermark removal, we used two residual operations to act the input and output of each Unet Block (block) for watermark removal. Due to the use of multiple concatenation operations in each U-net, two residual operations will result in the over-enhancement phenomenon of obtained features for image watermark removal. That is verified by both IWRU-net and IWRU-net with two extra residual operations in Table 1. To allow readers more easily to observe the effect of the IWRU-net, we chose one image from the LVW with one watermark randomly added onto the given clean image, rotated from -30° to 30° , scaled from 70% to 100% and adjusted to

a transparency of 50% to 80% to conduct the visual watermark removal on the image. As presented in Figure 2, we can see that our IWRU-net has obtained more clearly detailed information than that obtained by the IWRU-net with two extra residual operations in the observation area. This shows that two extra residual operations have a native effect on the IWRU-net for image removal. In other words, the designed serial architecture is capable to deal with watermark removal from images.

Table 1. PSNR (dB) results of different methods on 100 watermark images from the LVW for image watermark removal.

Methods	PSNR
IWRU-net (ours)	44.85
IWRU-net with four up- and down-sampling operations in each U-net	34.75
IWRU-net without up- and down-sampling operations	43.18
A single U-net	43.71
IWRU-net with two extra residual operations	36.77

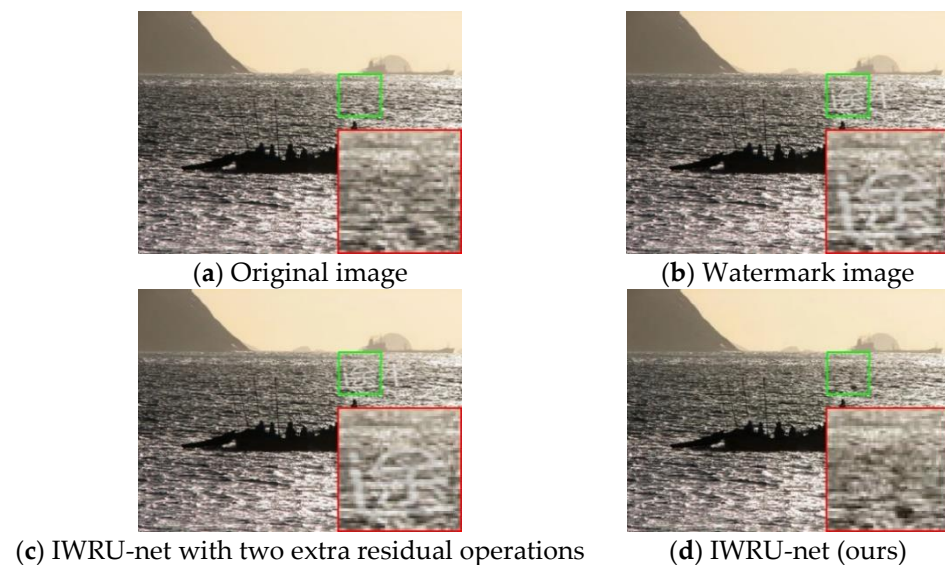


Figure 2. Visualization of different methods of watermark removal on an image from the LVW.

4.4. Experimental Results

Because image watermark removal is a low-level vision task, we chose a denoising convolutional neural network (DnCNN) [53], a fast and flexible denoising convolutional neural network (FFDNet) [56], a U-net [57], an attention-guided denoising convolutional neural network (ADNet) [58] and a robust deformed denoising CNN (RDDCNN) [31] as comparative methods to test the performance of image watermark removal in terms of qualitative and quantitative evaluations on the LVW and CLWD. For qualitative evaluation, we first used transparency rates of 1 and 0.5 to test the effects on the IWRU-net for image watermark removal. As shown in Table 2, our IWRU-net obtained a higher PSNR at a transparency of 1 than at 0.5 for image watermark removal, where one image is chosen from the LVW and its other setting is the same as in Section 4.2. This also shows that, when the transparency is lower, the IWRU-net has better results of watermark removal.

Table 2. PSNR (dB) results at different transparency rates of 100% and 50% with the IWRU-net for image watermark removal.

Transparency Rate	PSNR
100%	45.67
50%	41.32

Then, 100 randomly selected images from the LVW and CLWD datasets in Section 4.2 were used to test the effects of the transparency rate by varying it from 0.5, 0.6, 0.7 to 0.8 on the IWRU-net for image watermark removal. As illustrated in Tables 3 and 4, we can see that our IWRU-net is more effective than other methods, i.e., DnCNN, FFDNet and Unet for LVW and CLWD in terms of the PSNR and structural similarity (SSIM) [59] for blind watermark removal. This implies the robustness of our IWRU-net for image watermark removal.

Next, we also measured the complexity (parameters and flops [60]) of different methods, i.e., DnCNN, FFDNet, Unet, ADNet, RDDCNN and IWRU-net on an image with 512×512 from the LVW. Also, we used one image with 256×256 , 512×512 , and 1024×1024 from the LVW to test the running time of different methods, i.e., DnCNN, FFDNet, Unet, ADNet, RDDCNN and IWRU-net. As presented in Tables 5 and 6, we can see that our IWRU-net also his acceptably effective in terms of complexity and running time for image watermark removal. Compared with the cited methods, it is clear that our proposed IWRU-net is competitive for image watermark removal.

Table 3. Average PSNR (dB) and SSIM of different networks on LVW datasets for varying transparency rates of 0.5, 0.6, 0.7 and 0.8.

Methods	PSNR	SSIM
DnCNN [53]	42.95	0.9961
FFDNet [56]	38.48	0.9847
Unet [57]	43.71	0.9963
IWRU-net (ours)	44.85	0.9970

Table 4. Average PSNR (dB) and SSIM of different networks on CLWD datasets for varying transparency rates of 0.5, 0.6, 0.7 and 0.8.

Methods	PSNR	SSIM
DnCNN [53]	44.67	0.9753
FFDNet [56]	37.54	0.9912
Unet [57]	45.35	0.9972
RDDCNN [31]	46.25	0.9971
ADNet [58]	46.47	0.9972
IWRU-net (ours)	46.52	0.9975

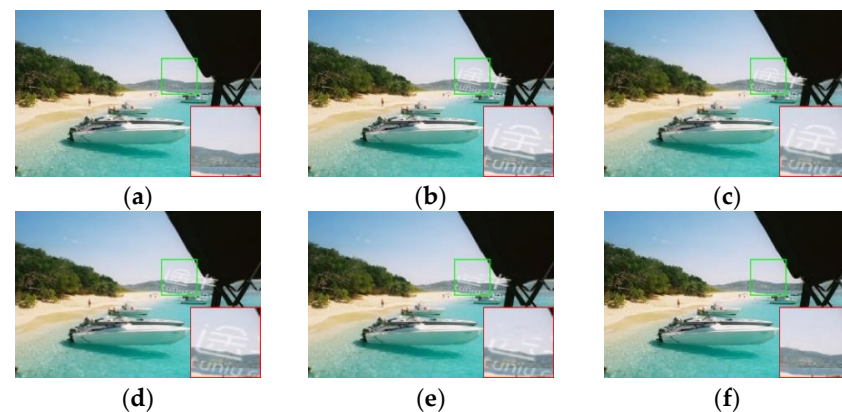
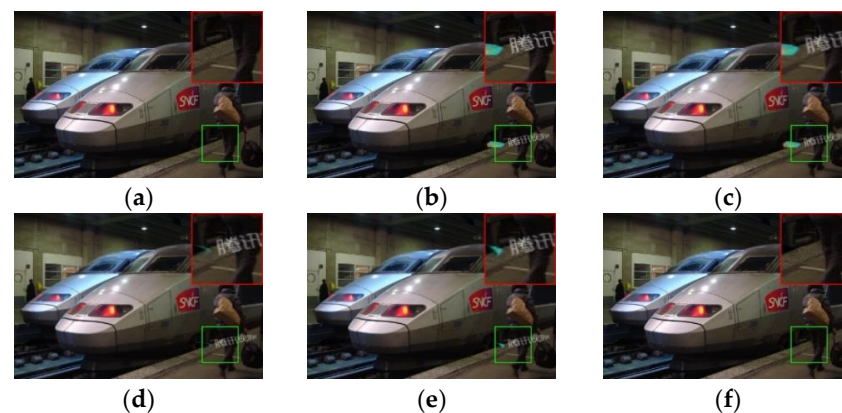
Table 5. The complexity of different watermark removal methods.

Methods	Parameters	Flops
DnCNN [53]	0.5594 M	36.6582 G
FFDNet [56]	0.4945 M	8.1023 G
Unet [57]	1.0120 M	18.6813 G
RDDCNN [31]	0.5591 M	36.7060 G
ADNet [58]	0.5215 M	34.2393 G
IWRU-net (ours)	2.0240 M	37.3625 G

Table 6. Running time for different watermark removal methods on three image sizes.

Methods	256×256	512×512	1024×1024
DnCNN [53]	0.038228	0.154801	0.638453
FFDNet [56]	0.010732	0.037471	0.124227
Unet [57]	0.027889	0.097742	0.316260
RDDCNN [31]	0.057355	0.222245	1.559665
ADNet [58]	0.036286	0.147691	0.563838
IWRU-net (ours)	0.058375	0.199374	0.654419

To further test the performance of our IWRU-net, we used quantitative evaluation to conduct visual effects as follows. We chose four images from the LVW, adding different transparency rates of 0.5, 0.6, 0.7 and 0.8, respectively, to test the visual effects of the IWRU-net. Also, DnCNN, FFDNet and Unet were used as comparative methods. One chosen area of each predicted visual image was enlarged as an observation area. The observation area is clearer, so its corresponding method has better performance in image watermark removal. As shown in Figures 3–6, we can see that our IWRU-net has clearer areas for different transparency rates. It shows that our IWRU-net is more advantageous in terms of quantitative evaluation for image watermark removal. According to these findings, it is known that the IWRU-net is very suitable to image watermark removal for qualitative and quantitative evaluations.

**Figure 3.** Visual effects of different methods with a transparency rate of 0.5 for image watermark removal. (a) Original image, (b) watermark image, (c) DnCNN, (d) FFDNet, (e) U-net and (f) IWRU-net (ours).**Figure 4.** Visual effects of different methods with a transparency rate of 0.6 for image watermark removal. (a) Original image, (b) watermark image, (c) DnCNN, (d) FFDNet, (e) U-net and (f) IWRU-net (ours).

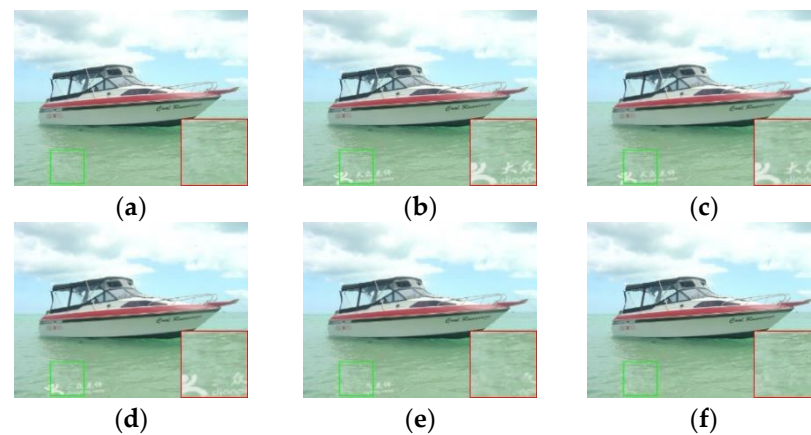


Figure 5. Visual effects of different methods with a transparency rate of 0.7 for image watermark removal. (a) Original image, (b) watermark image, (c) DnCNN, (d) FFDNet, (e) U-net and (f) IWRU-net (ours).



Figure 6. Visual effects of different methods with a transparency rate of 0.8 for image watermark removal. (a) Original image, (b) watermark image, (c) DnCNN, (d) FFDNet, (e) U-net and (f) IWRU-net (ours).

5. Conclusions

We propose an improved watermark removal U-net as IWRU-net. To improve the robustness of the obtained information, a serial architecture was used to facilitate more accurate information to guarantee the performance for watermark removal. To address long-term dependency problems, U-nets as simple components were integrated into the serial architecture to extract more salient hierarchical information for addressing watermark removal problems. To increase the adaptability of IWRU-net on mobile devices in the real world, randomly distributed watermarks of different types were used to train a blind watermark removal model. Our method is competitive with other popular watermark removal methods in terms of quantitative and qualitative evaluations. In the future, we will design lightweight CNNs for image watermark removal.

Author Contributions: Validation and part idea, L.F.; Data curation, writing and validation, B.S.; Investigation, L.S.; Part analysis, J.Z.; Part idea, visualization, writing (review), D.C.; Editing, H.Z.; Writing and Funding acquisition, C.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2021A1515110079, in part by the Fundamental Research Funds for the Central Universities under Grant D5000210966.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Swanson, M.D.; Zhu, B.; Tewfik, A.H. Transparent robust image watermarking. In Proceedings of the 3rd IEEE International Conference on Image Processing, Lausanne, Switzerland, 19 September 1996; IEEE: Piscataway, NJ, USA, 1996; Volume 3, pp. 211–214.
- Wong, P.W. A public key watermark for image verification and authentication. In Proceedings of the 1998 International Conference on Image Processing, ICIP98 (Cat. No. 98CB36269), Chicago, IL, USA, 7 October 1998; IEEE: Piscataway, NJ, USA, 1998; Volume 1, pp. 455–459.
- Park, J.; Tai, Y.W.; Kweon, I.S. Identigram/watermark removal using cross-channel correlation. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 446–453.
- Hsu, T.C.; Hsieh, W.S.; Chiang, J.Y.; Su, T. New watermark-removal method based on Eigen-image energy. *IET Inf. Secur.* **2011**, *5*, 43–50. [\[CrossRef\]](#)
- Boyle, R.D.; Hiary, H. Watermark location via back-lighting and recto removal. *Int. J. Doc. Anal. Recognit. (IJDAR)* **2009**, *12*, 33–46. [\[CrossRef\]](#)
- Yang, Y.; Sun, X.; Yang, H.; Li, C. Removable visible image watermarking algorithm in the discrete cosine transform domain. *J. Electron. Imaging* **2008**, *17*, 033008. [\[CrossRef\]](#)
- Makbol, N.M.; Khoo, B.E.; Rassem, T.H. Block-based discrete wavelet transform-singular value decomposition image watermarking scheme using human visual system characteristics. *IET Image Process.* **2016**, *10*, 34–52. [\[CrossRef\]](#)
- Ansari, I.A.; Pant, M. Multipurpose image watermarking in the domain of DWT based on SVD and ABC. *Pattern Recognit. Lett.* **2017**, *94*, 228–236. [\[CrossRef\]](#)
- Huynh-The, T.; Banos, O.; Lee, S.; Yoon, Y.; Le-Tien, T. Improving digital image watermarking by means of optimal channel selection. *Expert Syst. Appl.* **2016**, *62*, 177–189. [\[CrossRef\]](#)
- Fares, K.; Amine, K.; Salah, E. A robust blind color image watermarking based on Fourier transform domain. *Optik* **2020**, *208*, 164562. [\[CrossRef\]](#)
- Huang, Y.; Niu, B.; Guan, H.; Zhang, S. Enhancing image watermarking with adaptive embedding parameter and PSNR guarantee. *IEEE Trans. Multimed.* **2019**, *21*, 2447–2460. [\[CrossRef\]](#)
- Chen, X.; Wang, W.; Ding, Y.; Bender, C.; Jia, R.; Li, B.; Song, D.X. Leveraging unlabeled data for watermark removal of deep neural networks. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 10–15 June 2019; pp. 1–6.
- Sharma, S.S.; Chandrasekaran, V. A robust hybrid digital watermarking technique against a powerful CNN-based adversarial attack. *Multimed. Tools Appl.* **2020**, *79*, 32769–32790. [\[CrossRef\]](#)
- Haribabu, K.; Subrahmanyam, G.; Mishra, D. A robust digital image watermarking technique using auto encoder based convolutional neural networks. In Proceedings of the 2015 IEEE Workshop on Computational Intelligence: Theories, Applications and Future Directions (WCI), Kanpur, India, 14–17 December 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 1–6.
- Chen, X.; Wang, W.; Bender, C.; Ding, Y.; Jia, R.; Li, B.; Song, D.X. Refit: A unified watermark removal framework for deep learning systems with limited data. In Proceedings of the 2021 ACM Asia Conference on Computer and Communications Security, Hong Kong, China, 7–11 June 2021; pp. 321–335.
- Liu, Y.; Zhu, Z.; Bai, X. Wdnet: Watermark-decomposition network for visible watermark removal. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 4–8 January 2022; pp. 3685–3693.
- Lu, J.; Ni, J.; Su, W.; Xie, H. Wavelet-Based CNN for Robust and High-Capacity Image Watermarking. In Proceedings of the 2022 IEEE International Conference on Multimedia and Expo (ICME), Taipei, Taiwan, 18–22 July 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–6.
- Cao, L.; Liang, Y.; Lv, W.; Park, K.; Miura, Y.; Shinomiya, Y.; Yoshida, S. Relating brain structure images to personality characteristics using 3D convolution neural network. *CAAI Trans. Intell. Technol.* **2021**, *6*, 338–346. [\[CrossRef\]](#)
- Jafarbigloo, S.K.; Danyali, H. Nuclear atypia grading in breast cancer histopathological images based on CNN feature extraction and LSTM classification. *CAAI Trans. Intell. Technol.* **2021**, *6*, 426–439. [\[CrossRef\]](#)
- Cheng, D.; Li, X.; Li, W.; Lu, C.; Li, F.; Zhao, H.; Zheng, W. Large-scale visible watermark detection and removal with deep convolutional networks. In Proceedings of the Chinese Conference on Pattern Recognition and Computer Vision (PRCV), Guangzhou, China, 23–26 November 2018; Springer: Cham, Switzerland, 2018; pp. 27–40.
- Lee, J.E.; Seo, Y.H.; Kim, D.W. Convolutional neural network-based digital image watermarking adaptive to the resolution of image and watermark. *Appl. Sci.* **2020**, *10*, 6854. [\[CrossRef\]](#)
- Li, T.; Feng, B.; Li, G.; Li, X.; He, M.; Li, P. Visible Watermark Removal Based on Dual-input Network. In Proceedings of the 2021 ACM International Conference on Intelligent Computing and its Emerging Applications, Jinan, China, 28–29 December 2021; pp. 46–52.
- Meng, Z.; Morizumi, T.; Miyata, S.; Kinoshita, H. An Improved Design Scheme for Perceptual Hashing based on CNN for Digital Watermarking. In Proceedings of the 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC), Madrid, Spain, 13–17 July 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1789–1794.
- Le Merrer, E.; Perez, P.; Trédan, G. Adversarial frontier stitching for remote neural network watermarking. *Neural Comput. Appl.* **2020**, *32*, 9233–9244. [\[CrossRef\]](#)
- Ingaleswar, S.; Dharwadkar, N.V. Water chaotic fruit fly optimization-based deep convolutional neural network for image watermarking using wavelet transform. In *Multimedia Tools and Applications*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 1–25.

26. Li, Q.; Wang, X.; Ma, B.; Wang, X.; Wang, C.; Gao, S.; Shi, Y. Concealed attack for robust watermarking based on generative model and perceptual loss. In *IEEE Transactions on Circuits and Systems for Video Technology*; IEEE: Piscataway, NJ, USA, 2021.
27. Dhaya, R. Light weight CNN based robust image watermarking scheme for security. *J. Inf. Technol. Digit. World* **2021**, *3*, 118–132.
28. Chacko, A.; Chacko, S. Deep learning-based robust medical image watermarking exploiting DCT and Harris hawks optimization. *Int. J. Intell. Syst.* **2022**, *37*, 4810–4844. [[CrossRef](#)]
29. Wang, C.; Hao, Q.; Xu, S.; Ma, B.; Xia, Z.; Li, Q.; Li, J.; Shi, Y.Q. RD-IWAN: Residual Dense based Imperceptible Watermark Attack Network. In *IEEE Transactions on Circuits and Systems for Video Technology*; IEEE: Piscataway, NJ, USA, 2022.
30. Qin, H.; Yan, J.; Li, X.; Hu, X. Joint training of cascaded CNN for face detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 1–26 July 2016; pp. 3456–3465.
31. Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; Zhang, L. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Trans. Image Process.* **2017**, *26*, 3142–3155. [[CrossRef](#)] [[PubMed](#)]
32. Schlemper, J.; Caballero, J.; Hajnal, J.V.; Price, A.N.; Rueckert, D. A deep cascade of convolutional neural networks for MR image reconstruction. In Proceedings of the International Conference on Information Processing in Medical Imaging, Boone, CA, USA, 25–30 June 2017; Springer: Cham, Switzerland, 2017; pp. 647–658.
33. Wu, D.; Kim, K.; Fakhri, G.E.; Li, Q. A cascaded convolutional neural network for X-ray low-dose CT image denoising. *arXiv* **2017**, arXiv:1705.04267.
34. Li, C.; Guo, J.; Porikli, F.; Fu, H.; Pang, Y. A cascaded convolutional neural network for single image dehazing. *IEEE Access* **2018**, *6*, 24877–24887. [[CrossRef](#)]
35. Yan, S.; Wu, C.; Wang, L.; Xu, F.; An, L.; Guo, K.; Liu, Y. Ddrnet: Depth map denoising and refinement for consumer depth cameras using cascaded cnns. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 151–167.
36. Zhao, S.; Dong, Y.; Chang, E.I.; Xu, Y. Recursive cascaded networks for unsupervised medical imageregistration. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 10600–10610.
37. Hang, R.; Liu, Q.; Hong, D.; Ghamisi, P. Cascaded recurrent neural networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5384–5394. [[CrossRef](#)]
38. Wu, J.; Ma, J.; Liang, F.; Dong, W.; Shi, G.; Lin, W. End-to-end blind image quality prediction with cascaded deep neural network. *IEEE Trans. Image Process.* **2020**, *29*, 7414–7426. [[CrossRef](#)]
39. Tian, C.; Xu, Y.; Zuo, W.; Zhang, B.; Fei, L.; Lin, C. Coarse-to-fine CNN for image super-resolution. *IEEE Trans. Multimed.* **2020**, *23*, 1489–1502. [[CrossRef](#)]
40. Lu, X.; Zhang, J.; Yang, D.; Xu, L.; Jia, F. Cascaded convolutional neural network-based hyperspectral image resolution enhancement via an auxiliary panchromatic image. *IEEE Trans. Image Process.* **2021**, *30*, 6815–6828. [[CrossRef](#)] [[PubMed](#)]
41. Xue, F.; Tan, Z.; Zhu, Y.; Ma, Z.; Guo, G. Coarse-to-fine cascaded networks with smooth predicting for video facial expression recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 4 November 2022; pp. 2412–2418.
42. Tian, C.; Yuan, Y.; Zhang, S.; Lin, C.; Zuo, W.; Zhang, D. Image Super-resolution with An Enhanced Group Convolutional Neural Network. *arXiv* **2022**, arXiv:2205.14548. [[CrossRef](#)] [[PubMed](#)]
43. Tian, C.; Zhang, Y.; Zuo, W.; Lin, C.; Zhang, D.; Yuan, Y. A heterogeneous group CNN for image super-resolution. *arXiv* **2022**, arXiv:2209.12406. [[CrossRef](#)] [[PubMed](#)]
44. Bloomfield, P.; Steiger, W.L. *Least Absolute Deviations: Theory, Applications, and Algorithms*; Birkhäuser: Boston, MA, USA, 1983.
45. Pollard, D. Asymptotics for least absolute deviation regression estimators. *Econom. Theory* **1991**, *7*, 186–199. [[CrossRef](#)]
46. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
47. Agarap, A.F. Deep learning using rectified linear units (relu). *arXiv* **2018**, arXiv:1803.08375.
48. Murray, N.; Perronnin, F. Generalized max pooling. In Proceedings of the IEEE conference on computer vision and pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 2473–2480.
49. Xu, J.; Li, Z.; Du, B.; Zhang, M.; Liu, J. Reluplex made more practical: Leaky ReLU. In Proceedings of the 2020 IEEE Symposium on Computers and communications (ISCC), Rennes, France, 7–10 July 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–7.
50. Li, X.; Lu, C.; Cheng, D.; Li, W.; Cao, M.; Liu, B.; Ma, J.; Zheng, W. Towards photo-realistic visible watermark removal with conditional generative adversarial networks. In Proceedings of the International Conference on Image and Graphics, Beijing, China, 23–25 August 2019; Springer: Cham, Switzerland, 2019; pp. 345–356.
51. Liang, J.; Niu, L.; Guo, F.; Long, T.; Zhang, L. Visible Watermark Removal via Self-calibrated Localization and Background Refinement. In Proceedings of the 29th ACM International Conference on Multimedia, Chengdu, China, 20 October 2021; pp. 4426–4434.
52. Cun, X.; Pun, C.M. Split then refine: Stacked attention-guided ResUNets for blind single image visible watermark removal. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtually, 2–9 February 2021; Volume 35, pp. 1184–1192.
53. Zhang, Q.; Xiao, J.; Tian, C.; Chun Wei Lin, J.; Zhang, S. A robust deformed convolutional neural network (CNN) for image denoising. *CAAI Trans. Intell. Technol.* **2022**. [[CrossRef](#)]
54. Khan, A.; Sohail, A.; Zahoora, U.; Qureshi, A.S. A survey of the recent architectures of deep convolutional neural networks. *Artif. Intell. Rev.* **2020**, *53*, 5455–5516. [[CrossRef](#)]
55. Hore, A.; Ziou, D. Image quality metrics: PSNR vs. SSIM. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 2366–2369.

-
56. Zhang, K.; Zuo, W.; Zhang, L. FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE Trans. Image Process.* **2018**, *27*, 4608–4622. [[CrossRef](#)]
 57. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; Springer: Cham, Switzerland, 2015; pp. 234–241.
 58. Tian, C.; Xu, Y.; Li, Z.; Zuo, W.; Fei, L.; Liu, H. Attention-guided CNN for image denoising. *Neural Netw.* **2020**, *124*, 117–129. [[CrossRef](#)] [[PubMed](#)]
 59. Setiadi, D.R.I.M. PSNR vs SSIM: Imperceptibility quality assessment for image steganography. *Multimed. Tools Appl.* **2021**, *80*, 8423–8444. [[CrossRef](#)]
 60. Dolbeau, R. Theoretical peak FLOPS per instruction set: A tutorial. *J. Supercomput.* **2018**, *74*, 1341–1377. [[CrossRef](#)]