Deep Inverse Halftoning via Progressively Residual Learning

Menghan Xia[2,1] and Tien-Tsin Wong[1,2] ⋆

[1] The Chinese University of Hong Kong, Sha Tin, Hong Kong
[2] Shenzhen Key Laboratory of Virtual Reality and Human Interaction Technology, SIAT, Shenzhen, China
{mhxia, ttwong}@cse.cuhk.edu.hk

Abstract. Inverse halftoning as a classic problem has been investigated in the last two decades, however, it is still a challenge to recover the continuous version with accurate details from halftone images. In this paper, we present a statistic learning based method to address it, leveraging Convolutional Neural Network(CNN) as a nonlinear mapping function. To exploit features as completely as possible, we propose a Progressively Residual Learning (PRL) network that synthesizes the global tone and subtle details from the halftone images in a progressive manner. Particularly, it contains two modules: Content Aggregation that removes the halftone patterns and reconstructs the continuous tone firstly, and Detail Enhancement that boosts the subtle structures incrementally via learning a residual image. Benefiting from this efficient architecture, the proposed network is superior to all the candidate networks employed in our experiments for inverse halftoning. Also, our approach outperforms the state of the art with a large margin.

Keywords: Inverse Halftoning · Progressive Learning

1 Introduction

Halftoning is the rendering of continuous-tone pictures on media where only two levels can be displayed. The problem arose in the late 19th century when printing machines attempted to print images on paper. This was accomplished by adjusting the size of the dots according to local image intensity, called analog halftoning. And digital halftoning, as a kind of image processing technique, transforms a continuous-tone image with discrete gray levels into a bi-level halftone image. It has been used in bi-level output devices such as printers, copiers, fax machines, and even in plasma display panels, for example, the inkjet printers use halftone image to determine the spatial position of the ink that drops on paper by the heating and localized vaporization of liquid in a jet chamber [14]. The commonly used halftoning methods are the clustered-dot dithering, blue-noise dithering, error diffusion, and direct binary search [32,21,4,18].

Inverse halftoning, the counter process of digital halftoning, reconstructs a continuous-tone image with 255 levels or more levels from its halftoned

⋆ Corresponding author: Tien-Tsin Wong

version [23]. As a kind of important image restoration, there are a wide range of applications of inverse halftoning. First, working with effective inverse halftoning techniques, digital halftoning might be used as an image compression method by representing images with lower bits. In addition, photos printed on books, magazines showing halftone patterns can be scanned and transformed into a continuous-tone image, which is especially meaningful for those historically important photos on old newspapers. Further, continuous-tone images also need to be reconstructed for image manipulations, such as sharpening, resizing, rotation, or tone correction to obtain much better results [29]. However, due to the inevitable information loss during the halftoning process, inverse halftoning of image is an ill-posed problem. Anyway, it is still possible to approximate a continuous-tone image from its bi-level origins since the uniform dot patterns could give a visual impression of gray. This indicates that a simple low-pass filter is applicable to halftoning inverse, but the results by such a simple operation are prone to blur and artifacts. Thus, one reasonable solution for inverse halftoning is to provide enough prior information to guide the reconstruction of continuous-tone images, which is just the reason why the look up table (LUT) based approach [19] has been the state-of-the-art among all the conventional methods. However, this heuristic strategy that continuous-tone patch is used to replace the halftone patch based on the neighboring information can only achieve results of limited quality.

Inspired by the powerful representation ability of CNN, we propose to address this problem through training a CNN model as a nonlinear function in supervised manner. That is to say, by learning the corresponding pairs of halftone and continuous-tone images, the trained CNN would be embedded with prior knowledge and be able to reconstruct any continuous-tone image from its halftone input because of its generalization ability. Generally, recovering the continuous tone from a halftone image involves two tasks: the one is to synthesize the global tone, and the other is to extract the structural details. Considering this nature, we specifically design a PRL network that consists of two corresponding modules: Content Aggregation and Detail Enhancement. Under an end-to-end training scheme, the content aggregation module synthesizes an initial continuous-tone images firstly, and the detail enhancement module further extracts fine structures by learning a residual image. Convincing experimental results illustrates the superiority of PRL over several candidate networks with typical architectures, especially in recovering subtle structures. In addition, running on the inverse halftoning benchmark, our approach set a new state-of-the-art performance.

2 Related Works

As our approach exploits CNN to address the classic problem of inverse halftoning, we review on the conventional inverse halftoning algorithms and typical/relevant CNN works.

2.1 Conventional Inverse Halftoning

Over the last decades, a variety of inverse halftoning methods have been proposed. Although a simple low-pass filtering can remove most of the halftoning noise but it also removes edge information. Thus, most of works have been focused on developing an edge preserving filter while suppressing the visible dot patterns. Given the prior knowledge about a specific halftone pattern, customized filters are proposed to compute continuous grayscale values [35,33] based on neighboring dot patterns. Without requiring prior information, Analoui et al. [22] propose a method of projection onto convex sets for halftone images produced by ordered dithering. This method has also been successfully applied for the tone restoration of error diffused halftones [30]. Another method for inverse halftoning of error diffused images is introduced by Kite et al. [31], which exploits space varying filtering based on gradients on the halftone images. Leveraging the overcomplete wavelet expansions, Xiong et al. [36] propose to separate the halftoning noises from the original image through edge detection, which outperforms all the other competing methods at that time.

Different from those aforementioned methods, Mese et al. [19] creatively propose to precompute a look up table (LUT) first and all the continuous-tone values are assigned depending on the neighborhood distribution of pixels in template. As reported in [20], the LUT based method is extremely fast and has a reconstructed quality comparable to the state-of-the-art methods of that time. This is reasonable since plenty of prior statistic information is provided for inverse halftoning in LUT. Attracted by this merit, most of recently proposed methods have taken the strategy of dictionary learning for inverse halftoning [28,29,25].

2.2 Convolutional Neural Networks

Recent deep CNNs have become a common workhorse behind a wide variety of image transformation problems. These problems can be formulated as per-pixel classification or regression by defining low level loss. Semantic segmentation methods [10,3,7] use fully convolutional neural networks trained by per-pixel classification loss to predict dense scene labels. End-to-end automatic image colorization techniques [26,5] try to colorize grayscale image based on low level losses. Other works for depth [2] and edge detection [34] are also similar to transform input images to meaningful output images through CNN, which are trained with per-pixel classification or regression loss. Denoise [11] and super-resolution [9,1] networks use CNNs to remove noise and generate fine structures respectively, which can be roughly regarded as two subtasks of inverse halftoning. Thus, some useful architecture elements of such networks might be used as reference, such as residual learning. Besides, most networks for image transformation use high-level feature spaces defined by pretrained CNN or generative adversarial networks (GAN) [8] to guarantee their output within the photo-realistic domain.

The outstanding ability of representing complex mappings makes deep CNN popular in various application. However, many works show that they are difficult to train because of the notorious vanishing gradient problem. To address problem, He et al. [12] propose the concept of residual blocks that use an "identity shortcut connection" to relieves the vanishing gradient issue. Models based on residual blocks have shown impressive performance in generative networks. However, some other works for image generation tasks [16,24] find that the batch normalization (BN) [17] used in residual block impedes the flexibility of the network, and they use the residual block with BN removed. These experimental conclusions are well absorbed in our network architecture.

3 Methodology

We formulate the inverse halftoning as image transformation problems and employ deep CNN as a nonlinear function to map halftone images to their continuous-tone version. The network architecture and loss function are presented in Section 3.1 and Section 3.2 respectively.

3.1 Network Architecture

As shown in Fig. 1, the proposed PRL network consists of two modules: content aggregation and detail enhancement. The content aggregation module begins with a normal convolutional layer and a residual block [12], followed by two downscaling blocks to spatially compress and encode the global information of the image. Then four residual blocks [12] with identical layout are used to construct the content and tone features, followed by two upscaling blocks bring the feature maps back to the input resolution. In detail enhancement module, we sequentially lay eight residual blocks to extract structural details

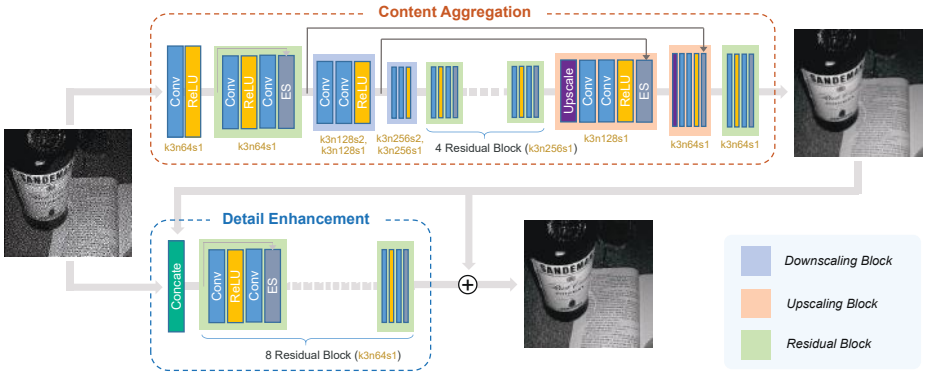

Fig. 1. Architecture of the proposed PRL network, where k is the kernel size, n is the number of feature maps and s is the stride in each convolutional layer, and ES denotes elementwise adding layer for feature maps.

Fig. 2. Output illustration of each module in the PRL network. For quantitative reference, the PSNR/SSIM against the groundtruth are annotated along.

in an incremental way. Such a simple yet effective strategy of using residual blocks is firstly adopted in the state-of-the-art super-resolution network [1] which majors in extracting fine structures from low-resolution images. Besides, providing auxiliary contextual cue, the output of content aggregation module is fed into the detail enhancement module along with the halftone input through a concatenation layer. Some readers might notice that the residual blocks used in our network use no batch normalization, which benefits in maintaining a higher flexibility of CNN [16,24]. The detailed parameters of each convolutional layer are annotated aside them in Figure 1.

Residual learning framework facilitates the learning efficiency of CNN. However, we can not employ the normal residual learning architecture directly because of the large gap between the input and the output in inverse halftoning problem. The key idea of PRL is to first generate an initial result which is roughly the same as the groundtruth and then focus on extracting those subtle information by learning a residual image of the initial result and the groundtruth. For inverse halftoning, the first task is relatively easier, and even a simple CNN with several layers is enough to remove halftone noise. Nonetheless, a more advanced architecture contributes to generating more details. The real challenge lies in how to recover those subtle structures from halftone patterns as accurate as possible. In our PRL network, an initial continuous-tone image is firstly synthesized by the content aggregation module, and then the residual image between the initial image and the groundtruth is learned incrementally by the detail enhancement module. Through this way, much richer details could be recovered on the final reconstructed continuous-tone image that is the sum of the initial image and the optimized residual image. The output examples of each component are illustrated in Fig. 2. As we can observe, the residual image mainly contains tiny boundaries and subtle textures. Even though the output with refined details only outperforms the initial image slightly in quantitative metrics, those further enhanced structural and texture details play an very importance role in contributing to more photo-realistic results.

3.2 Loss Function

To strengthen the feature extraction efficiency of CNN, we provide supervising information for the output of both content aggregation module G and the full network F during training. Therefore, the loss function includes two parts of computation, which is denoted as:

$$\mathcal{L}(F) = \mathcal{L}_{content}(G) + \omega \mathcal{L}_{full}(F), \tag{1}$$

where ω is a weight to balance the two terms and $\omega = 1.5$ is used in all of our experiments. Obviously G is a part of F.

Let $\mathcal{P} = \{(\mathbf{I}_h^i, \mathbf{I}_c^i)\}_{i=1}^n$ be the whole training pairs of halftone image $\mathbf{I}_h$ and its continuous-tone version $\mathbf{I}_c$. $\mathcal{L}_{content}$ describes the difference between the predicted initial image $G(\mathbf{I}_h^i)$ and its ground truth $\mathbf{I}_c^i$. Considering this term only involves recovering the global information of the continuous-tone image, we employ the Mean Square Error (MSE) to compute its loss, expressed as:

$$\mathcal{L}_{content} = \mathbb{E}_{(\mathbf{I}_h^i, \mathbf{I}_c^i) \in \mathcal{P}} \{ ||G(\mathbf{I}_h^i) - \mathbf{I}_c^i||_2 \} \tag{2}$$

where $|| \bullet ||_2$ denotes the L_2 norm (MSE). $\mathbb{E}$ denotes to compute the average loss over all the training set $\mathcal{P}$. Although achieving particularly high PSNR, the optimization results under MSE loss often lack high-frequency details which shows perceptually unsatisfying effects with overly smooth textures. Therefore, to compute the loss of the final output $F(\mathbf{I}_h^i)$, we use a measure metric that is closer to perceptual similarity together with pixel-wise L_1 loss. To define the perception loss, we adopt the high-level feature maps of the pretrained VGG network [27] which has been demonstrated to have good object preservation ability. Accordingly, $\mathcal{L}_{full}$ is formulated in Eq. 3.

$$\mathcal{L}_{full}(F) = \mathbb{E}_{(\mathbf{I}_h^i, \mathbf{I}_c^i) \in \mathcal{P}} \{ \alpha ||VGG_l(F(\mathbf{I}_h^i)) - VGG_l(F(\mathbf{I}_c^i))||_2 + ||\mathbf{I}_h^i - \mathbf{I}_c^i||_1 \} \tag{3}$$

where $|| \bullet ||_1$ denotes the L_1 norm. l refers to the feature maps of a specific VGG layer. Alike to [15], $l = conv4\_4$ is used. $\alpha = 2.0 \times 10^{-6}$ is used to balance the magnitude difference between deep feature space and pixel value space.

3.3 Data Preparation and Training

There are lots of image halftoning algorithms, such as ordered dithering, error diffusion, etc. In this paper, we take the most popular one, Floyd Steinberg error diffusion, as an example for inverse halftoning experiments. Applying it to the others would the same story. Particularly, the publicly available VOC2012 Dataset[3] is used in experiment, of which over 13,000 images are randomly selected as our training images. We crop and resize them to 256×256 without aspect distortion as the continuous-tone groundtruth, and then employ the error diffusion algorithm on them to get the corresponding halftone images. Note that

[3] VOC2012: http://host.robots.ox.ac.uk/pascal/VOC/voc2012/

our network can process images of any size during testing phase since it is a fully convolutional network.

We implement our network with TensorFlow in Python language. All experiments are performed on an PC with one GPU of NVIDIA GeForce GTX 1080 Ti. We train our network end-to-end for 150 epochs. The learning rate is set as 0.0002 initially and linearly decreases to 0.0002/100 in the end. Adam [13] is employed to optimize the network parameters.

4 Results and Discussion

We perform experiments on two testing datasets, of which the one (*dataset-1*) contains 3000 images from VOC2012 dataset that has no overlap with the training part, the other (*dataset-2*) contains 5000 images that are randomly picked from PLACE205 Dataset[4]. For quantitative evaluation, we employ PSNR and SSIM to measure the similarity between the reconstructed images and the groundtruth. For color image, they are calculated in every channel of RGB color space independently and the average values are used. However, as presented in [1] by mean opinion score testing, PSNR and SSIM fail to capture and accurately assess image quality with respect to the human visual system. The focus of this paper is the perceptual quality of the reconstructed continuous-tone images, so we recommend readers to pay more attention to the qualitative evaluation.

Particularly, to concentrate on the essential problem of inverse halftoning, the gray version of the two datasets are used for evaluation and analysis. The code is available at: `https://github.com/MenghanXia/InverseHalftoning`

Table 1. PSNR and SSIM obtained by the plain-architecture network and our PRL network.

Network	*dataset-1*		*dataset-2*	
	PSNR	SSIM	PSNR	SSIM
Plain architecture	29.676	0.8807	29.567	0.8912
PRL architecture	29.705	0.8797	29.567	0.8919

4.1 Effect of the PRL architecture

To show the superiority of the proposed PRL architecture, we compare it with a plain version of our network that sequentially joints the content aggregation module and detail enhancement module without residual learning scheme. As the comparative results shown in Fig. 3, the PRL network has an superiority over the plain one in generating more accurate structural lines and richer details. This can be explained by that residual learning can ease the difficulty of learning the transformation from the input to the output, because it only need to synthesize a residual image instead of the whole output. Since there is

[4] PLACE205: http://places.csail.mit.edu/index.html

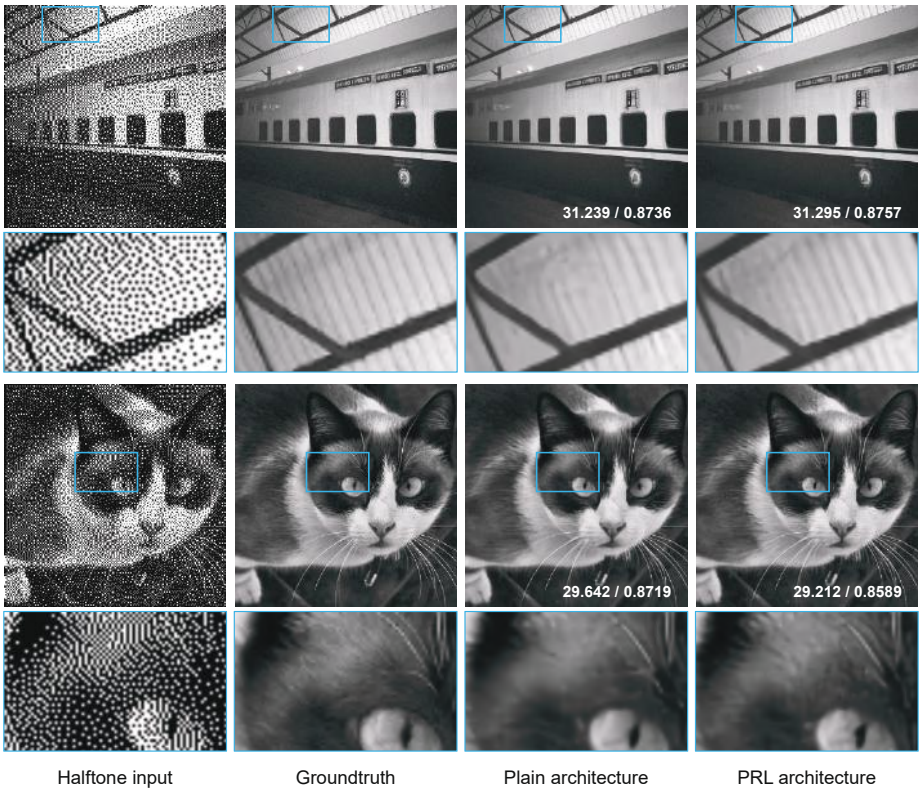

Fig. 3. Results of networks of plain architecture and the proposed PRL architecture. For detail observation, typical regions within blue box are enlarged below each image.

obvious gap between the input and the output in inverse halftoning problem, it is not reasonable to directly apply the normal residual learning architecture that is widely used in denoise network or super-resolution network. Instead we propose a progressively residual learning strategy that is especially suitable for the inverse halftoning task. Quantitative results are listed in Table 1. Even the superiority in PSNR and SSIM is minor, it dose have some visually observable benefits on the reconstructed images.

4.2 Comparative Evaluation

Comparison with state of the art: The latest inverse halftoning work [29] that exploits local learned dictionaries for edge optimization has been the state-of-the-art approach so far. Here, we compare it with our CNN based approach on six benchmark images. The benchmark images and the results of [29] are downloaded from the website provided on their paper. The comparative results are illustrated in Fig. 4, where the PSNR and SSIM of each image are also

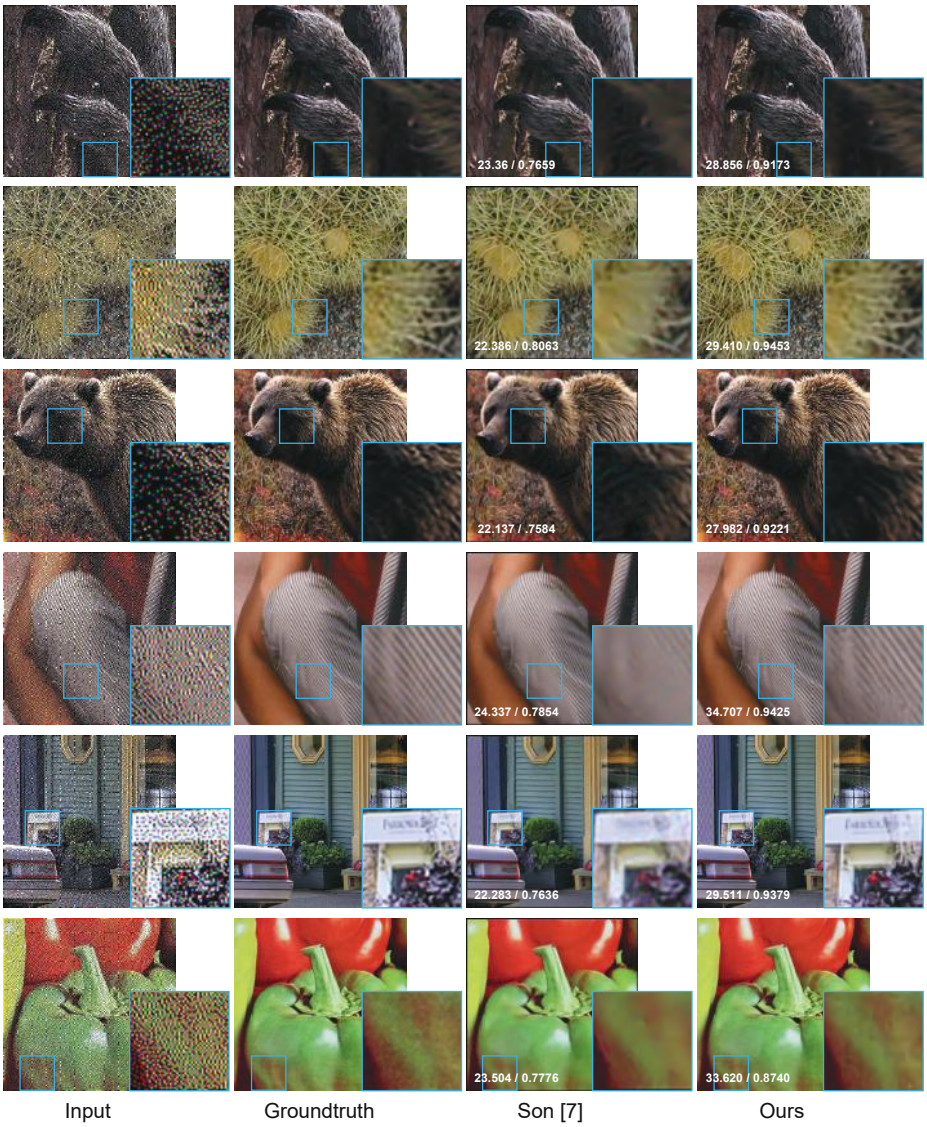

Fig. 4. Results of the state of the art [29] and our approach on the benchmark images. For quantitative comparison, PSNR and SSIM of each image are annotated along. The average values of [29] and ours are 23.002/0.7762 and 30.680/0.9232 respectively.

annotated for reference. As we can observe, our approach have a great superiority over the state of the art in both qualitative and quantitative evaluation. The results of [29] are overly smooth and lack of texture details, and some regions even shows residual halftone noise. On the contrary, the results of our approach present very close appearance to the groundtruth.

Table 2. Comparison on PSNR and SSIM obtained by different methods.

Method	*dataset-1*		*dataset-2*	
	PSNR	SSIM	PSNR	SSIM
DnCNN	**29.967**	0.8842	29.865	0.8961
VDSR	29.763	0.8822	29.791	0.8948
SRResNet	29.708	0.8763	29.608	0.8890
DeHalftone	29.404	0.8809	29.247	0.8929
Ours(mse)	29.914	**0.8854**	**29.915**	**0.8970**
Ours	29.705	0.8797	29.567	0.8919

Comparison with candidate networks: It seems a matter of course for a CNN based method to outperform a dictionaries learning based conventional method. We further compare our approach with other CNN based state-of-the-art methods of relevant tasks, like denoise and super resolution, which are extended for inverse halftoning here. DnCNN [11] exploits the residual learning idea for image noise removal. VDSR [9] employs a network with residual learning architecture for super resolution. SRResNet [1] integrates a seres of residual block sequentially in their super-resolution network, which is widely referenced in image transformation networks. Note that the upscaling layers of VDSR and SRResNet are removed in order to make them fit to our task. Besides, we also compare our approach with an unpublished method of inverse halftoning (DeHalftone) [6] that combines a normal U-Net architecture with perception loss. Due the powerful representation ability of CNN, all these methods have achieved high PSNR and SSIM on *dataset-1* and *dataset-2*, as illustrated in Table 2. Moreover, the quantitative differences is quite small, which might be explained by that all the methods are flexible enough to cover the transformation from halftone to continuous tone but their performances are just limited to the information capacity carried by the halftone input. Anyway, DnCNN and VDSR have the relatively higher PSNR and SSIM among all the methods, because they only employ MSE loss that conforms to the computation of PSNR and SSIM. We also provide the result of our approach with using MSE loss only, which shows improved PSNR and SSIM as well.

However, the higher PSNR and SSIM do not necessarily means the better visual quality, since the MSE loss benefits to a higher PSNR and SSIM but tends to cause perceptually unsatisfying effects with overly smooth textures. As illustrated in Fig. 7 and Fig. 8, we present comparatively qualitative results of difference methods on representative images from *dataset-1* and *dataset-2* respectively. Observing as a whole, the continuous-tone version of halftone images are recovered by all the methods equally in quality. When we take a close look at local details, as presented in those enlarged regions, some reconstruction differences among the results of different methods can be further observed, which are not well reflected by PSNR and SSIM. Thanks to the PRL architecture that works on both global and subtle features, our approach achieves the best reconstruction accuracy, especially in those slight but semantically important structure lines. This is an good example that illustrates the effectiveness of

residual learning. Anyway, for inverse halftoning, the PRL architecture has a better performance than the normal residual learning architecture that are used in DnCNN, VDSR and SRResNet.

4.3 Extended Applications

Providing corresponding training data, our method can be employed for several exciting applications. Ordered dithering is widely used for printing media, which transforms normal photos into halftone before printed on books or newspapers. Training our model with such halftone categories, we can transform those printed halftone photos back to their original continuous-tone version. This is especially meaningful for those historically important photos that only exist on printed materials now. In Fig. 5, we present two examples demonstrating this application. With a continuous-tone image, we can further apply all kinds of image editing and processing on them, while they are impossible for halftone images.

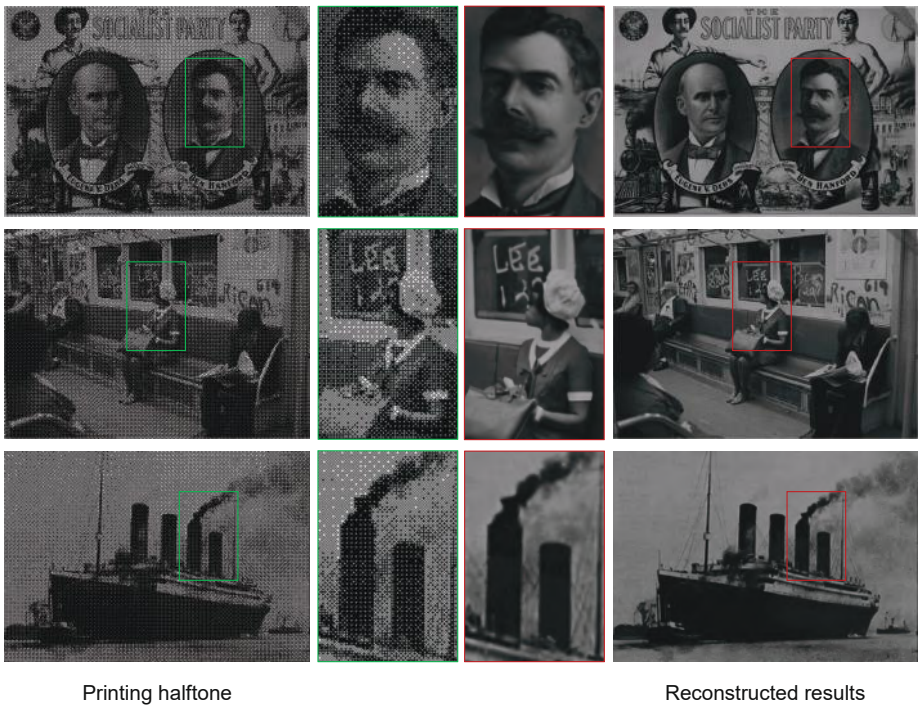

Fig. 5. Reconstruction results from printing halftone: ordered dithering.

Besides, storing a digital image in Graphics Interchange Format (GIF) also employs error diffusion, where only 256 colors are sampled to rendering the whole image based on its color distribution. Because of information loss, images of GIF often present annoying visual artifacts. Training our model with such halftone

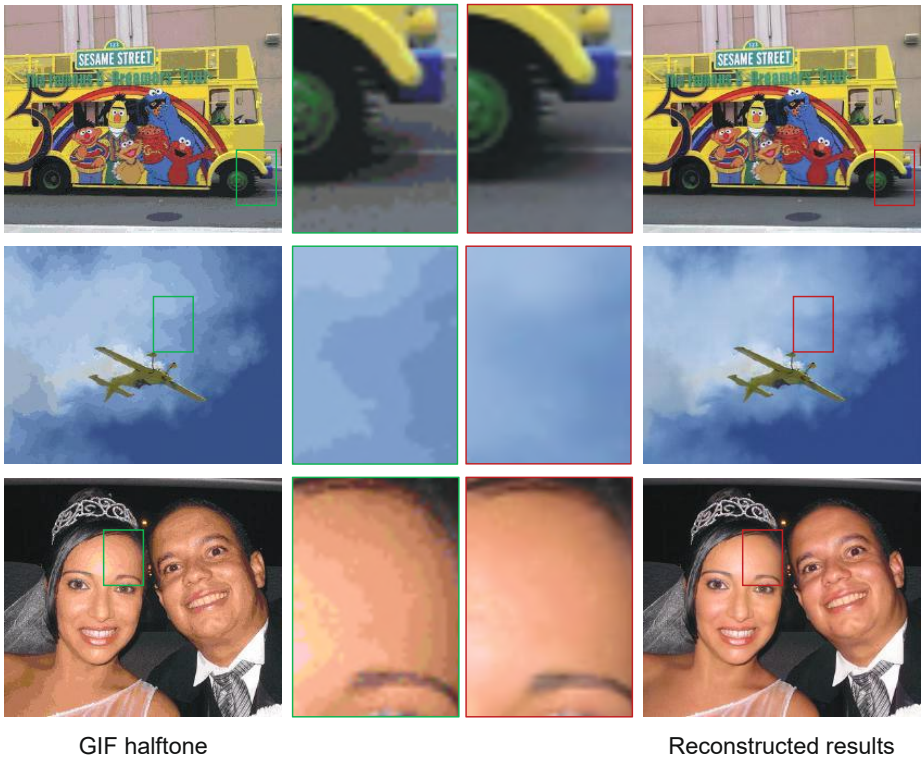

Fig. 6. Reconstruction results from GIF halftone: error diffusion.

images, we can transform such quantized images back to the visually pleasing continuous version. Two examples of this application are demonstrated in Fig. 6. In addition, we also provide a video demo displaying the dynamic effect of our processed results in the supplementary file.

5 Conclusion

We have proposed a PRL network that shows outstanding performance in accurately reconstructing continuous-tone images from halftone images. Running on the benchmark dataset, our approach outperforms the state of the art with a significant margin both qualitatively and quantitatively. In addition, comparing our network with other candidate networks shows that all the methods have obtained a very close PSNR and SSIM but they have significantly different performances on details recovery. It means the widely used PSNR and SSIM could not effectively reflect the perceptual quality of the results. Anyway, comparative results on enlarged regions illustrate the superiority of our approach in accurately recovering structural details. Besides, applying our approach to other halftone categories enables two practical applications successfully.

Acknowledgement: This project is supported by Shenzhen Science and Technology Program (No.JCYJ20160429190300857) and Shenzhen Key Laboratory (No.ZDSYS201605101739178), and the Research Grants Council of the Hong Kong Special Administrative Region, under RGC General Research Fund (Project No. CUHK14201017).

References

1. Christian Ledig, Lucas Theis, F.H.J.C.A.C.A.A.A.A.e.a.: Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
2. David Eigen, C.P., Fergus, R.: Depth map prediction from a single image using a multi-scale deep network. In: the Advances in Neural Information Processing Systems (NIPS) (2014)
3. Eigen, D., Fergus, R.: Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2015)
4. Floyd, R.: An adaptive algorithm for spatial gray-scale. In: Society of Information Display (1976)
5. Gustav Larsson, M.M., Shakhnarovich, G.: Learning representations for automatic colorization. In: European Conference on Computer Vision (ECCV) (2016)
6. Hou, X., Qiu, G.: Image companding and inverse halftoning using deep convolutional neural networks. arXiv preprint:1707.00116 (2017)
7. Hyeonwoo Noh, S.H., Han, B.: Learning deconvolution network for semantic segmentation. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2015)
8. Ian Goodfellow, Jean Pouget-Abadie, M.M.B.X.D.W.F.S.O.A.C., Bengio, Y.: Generative adversarial nets. In: Advances in Neural Information Processing Systems (NIPS) (2014)
9. Jiwon Kim, J.K.L., Lee, K.M.: Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
10. Jonathan Long, E.S., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015)
11. Kai Zhang, Wangmeng Zuo, Y.C.D.M., Zhang, L.: Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE Transactions on Image Processing (TIP) **26(7)**, 3142–3155 (2017)
12. Kaiming He, Xiangyu Zhang, S.R., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
13. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint:1511.06349 (2014)
14. Kipphan, H.: Handbook of print media. Springer Science & Business Media (2001)
15. Leon Gatys, A.S.E., Bethge, M.: Texture synthesis using convolutional neural networks. In: Advances in Neural Information Processing Systems (NIPS) (2015)

16. Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M.: Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (2017)
17. Loffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: IEEE International Conference on Machine Learning (ICML) (2015)
18. M.A.Seldowitz, J.A., Sweeney, D.: Synthesis of digital holograms by direct binary search. Applied Optics **26(14)**, 2788–2798 (1987)
19. Mese, M., Palghat, P.V.: Look-up table (lut) method for inverse halftoning. IEEE Transactions on Image Processing (TIP) **10(10)**, 1566–1578 (2001)
20. Mese, M., Vaidyanathan, P.P.: Recent advances in digital halftoning and inverse halftoning methods. IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications **49(6)**, 790–805 (2002)
21. Mitsa, T., Parker, K.: Digital halftoning technique using a blue-noise mask. JOSA A **9(11)**, 1920–1929 (1992)
22. Mostafa Analoui, J.A.: New results on reconstruction of continuous-tone from halftone. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASS) (1992)
23. Murat Mese, P.P.V.: Recent advances in digital halftoning and inverse halftoning methods. IEEE Transactions on Circuits and Systems **49(6)**, 790–806 (2002)
24. Nah, S., Kim, T.H., Lee, K.M.: Deep multi-scale convolutional neural network for dynamic scene deblurring. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
25. Pedro G. Freitas, M.C.F., Araujo, A.P.: Enhancing inverse halftoning via coupled dictionary training. Signal Processing: Image Communication **49**, 1–8 (2016)
26. Satoshi Lizuka, E.S.S., Ishikawa, H.: Let there be color!: joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. ACM Transactions on Graphics (TOG) **35(4)**, 110–121 (2016)
27. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition (2014), arXiv preprint:1409.1556
28. Son, C.H.: Inverse halftoning based on sparse representation. Optics letters **37(12)**, 2352–2354 (2012)
29. Son, C.H., Choo, H.: Local learned dictionaries optimized to edge orientation for inverse halftoning. IEEE Transactions on Image Processing (TIP) **23(6)**, 2542–2557 (2014)
30. Soren Hein, A.Z.: Halftone to continuous-tone conversion of error-diffusion coded images. Sigma Delta Modulators **213**, 133–154 (1993)
31. Thomas D. Kite, Damera-Venkata Niranjan, L.E.B., Alan, C.B.: A high quality, fast inverse halftoning algorithm for error diffused halftones. In: Proceedings of the IEEE International Conference on Image Processing (ICIP) (1998)
32. Ulichney, R.: Dithering with blue noise. In: Proceedings of the IEEE (1988)
33. Wong, P.W.: Inverse halftoning and kernel estimation for error diffusion. IEEE Transactions on Image Processing (TIP) **4(4)**, 486–498 (1995)
34. Xie, S., Tu, Z.: Holistically-nested edge detection. In: IEEE International Conference on Computer Vision (ICCV) (2015)
35. Yeong-Taeg Kim, R.A.G., Nikolai, G.: Inverse halftoning using binary permutation filters. IEEE Transactions on Image Processing (TIP) **4(9)**, 1296–1311 (1995)
36. Zixiang Xiong, T.O.M., Kannan, R.: Inverse halftoning using wavelets. IEEE Transactions on Image Processing (TIP) **8(10)**, 1479–1483 (1999)

Fig. 7. Qualitative comparisons among different methods on selected images of *dataset-1*. Typical region marked with blue box are enlarged for detail observation (red dash box indicates reconstruction of low accuracy). For quantitative reference, the PSNR and SSIM of each image are annotated.

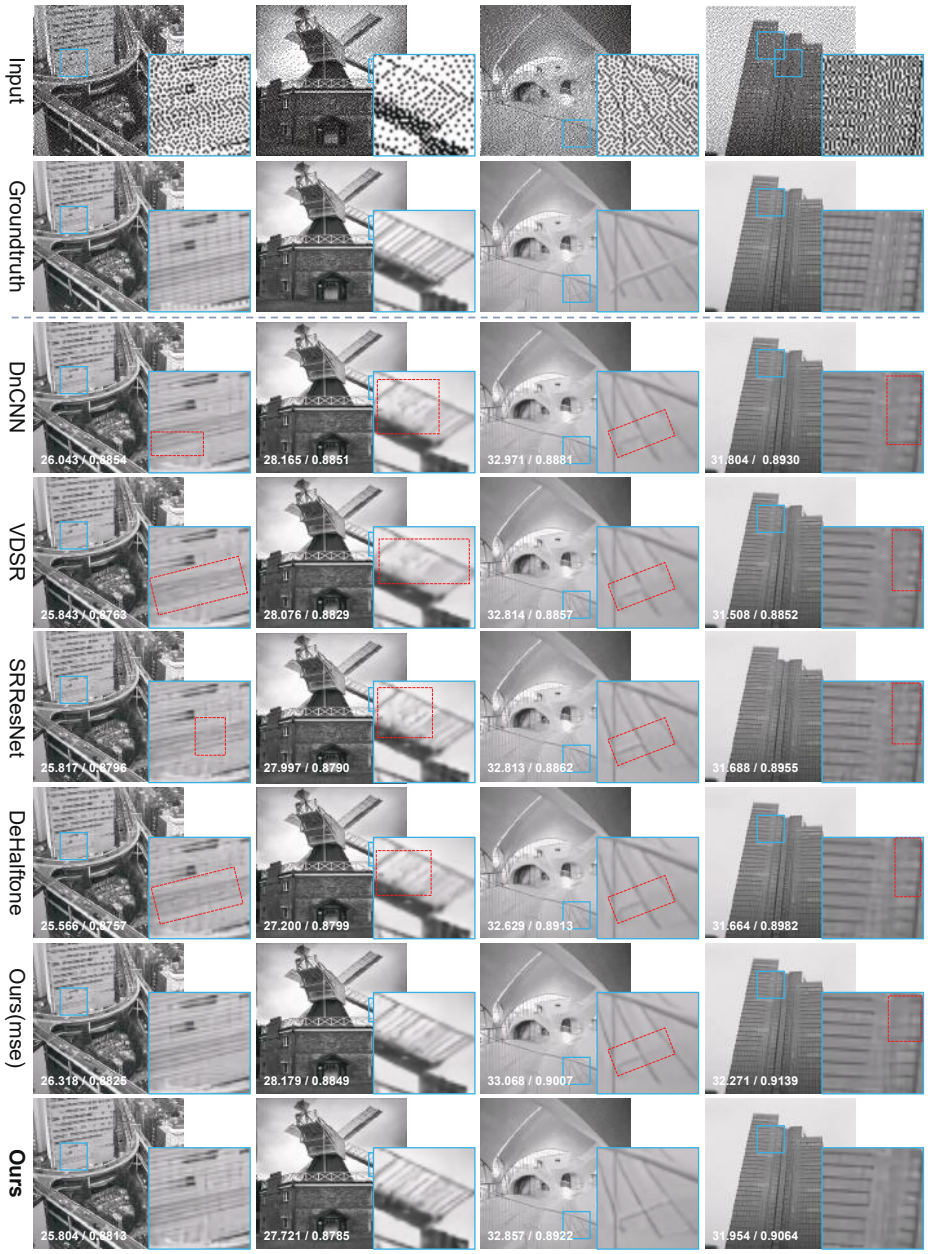

Fig. 8. Qualitative comparisons among different methods on selected images of *dataset-2*. Typical region marked with blue box are enlarged for detail observation (red dash box indicates reconstruction of low accuracy). For quantitative reference, the PSNR and SSIM of each image are annotated.