Supplementary Material-SDWNet: A Straight Dilated Network with Wavelet Transformation for image Deblurring

Wenbin Zou[1,*], Mingchao Jiang[2,*], Yunchen Zhang[3,*], Liang Chen[1,†], Zhiyong Lu[2], Yi Wu[1]
Fujian Provincial Key Laboratory of Photonics Technology, Fujian Normal University, Fuzhou, China.[1]
JOYY AI GROUP, Guangzhou, China.[2]
China Design Group Co., Ltd., Nanjing, China.[3]
alexzou14@foxmail.com, jiangshaoyu1993@gmail.com, cydiachen@cydiachen.tech,
cl_0827@126.com, yong1514@gmail.com, wuyi@fjnu.edu.cn

1. Introduction

This document includes supplementary information, which will be published as a PDF linked to the primary article. This supplementary document presents the following information that would be beneficial for the readers:

1. The principle of discrete wavelet transform
2. Additional ablation studies
   - Additional ablation on the dilated rate of DC module.
   - Additional ablation on the different upsampling methods.
   - Additional ablation on the effect of input size on the network.
3. The extended visual comparisons

2. The discrete wavelet transform

In practice, we usually use filter groups to implement the DWT. The input signal obtains the different frequency sub-bands through the corresponding high-pass filter $H(n)$ and low-pass filter $L(n)$. We take the most commonly used Haar wavelet [1] as an example, $L(n)$ and $H(n)$ of Haar wavelet can be defined as:

$$L(n) = \begin{cases} 1, & n = 0, 1 \\ 0, & \text{otherwise} \end{cases}, H(n) = \begin{cases} 1, & n = 0 \\ -1 & n = 1 \\ 0 & \text{otherwise} \end{cases}, \tag{1}$$

Suppose we have an image x with the size of $M \times N$, where M is the width, and N is the height. $X[m, n]$ represents the pixel value of the image at position (m, n), where $n = 1, 2, 3, \cdots, N$, $m = 1, 2, 3, \cdots, M$. 2D-DWT could be regarded as 1D-DWT is implemented in row $X[:, n]$ and column directions $X[m, :]$ successively. The 2D-DWT decomposes X into four frequency sub-bands, which could be written as LL, LH, HL, and HH. The process of single-level 2D-DWT for x could be represented as Figure. 1 and the following formula:

$$X_{1,L}[m, n] = \sum_{k=0}^{K-1} X[m, 2n - k]L_1[k], \tag{2}$$

$$X_{1,H}[m, n] = \sum_{k=0}^{K-1} X[m, 2n - k]H_1[k], \tag{3}$$

where X represents the input image, $X_{1,L}[m, n]$ and $X_{1,H}[m, n]$ are the results of 1D-DWT for each row.

$$X_{2,LL}[m, n] = \sum_{k=0}^{K-1} X_{1,L[m,n]}[2n - k, m]L_2[k], \tag{4}$$

$$X_{2,LH}[m, n] = \sum_{k=0}^{K-1} X_{1,L[m,n]}[2m - k, m]H_2[k], \tag{5}$$

$$X_{2,HL}[m, n] = \sum_{k=0}^{K-1} X_{1,H[m,n]}[2m - k, m]L_2[k], \tag{6}$$

$$X_{2,HH}[m, n] = \sum_{k=0}^{K-1} X_{1,H[m,n]}[2m - k, m]H_2[k], \tag{7}$$

Then, $X_{2,LL}[m, n]$, $X_{2,LH}[m, n]$, $X_{2,HL}[m, n]$, and $X_{2,HH}[m, n]$ are obtained by the 1D-DWT for each column.

*Equal contribution
†Corresponding author

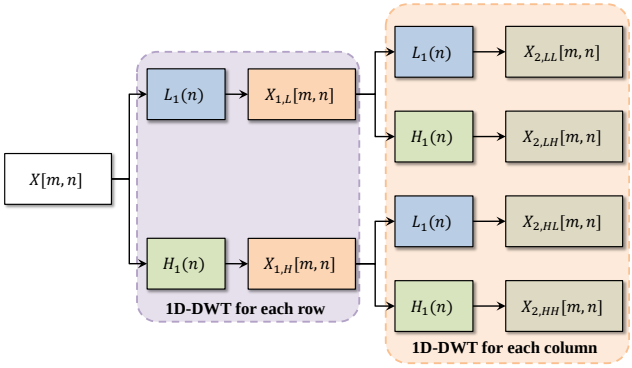

Figure 1. The processing of 2D-DWT.

Table 1. The effect of the different dilated rate of DC module on the network performance with an input of 96×96.

Dilated rate	PSNR	SSIM
$\{1, 2, 3, 4\}$	26.45	0.7490
$\{1, 2, 4, 8\}$	26.52	0.7564
$\{1, 3, 5, 7\}$	26.46	0.7503
$\{1, 3, 5, 7\}$, $\{1, 2, 4, 8\}$	27.22	0.7722
$\{1, 2, 4, 8\}$, $\{1, 3, 5, 7\}$	**27.35**	**0.7756**

3. Additional ablation studies

3.1. The dilated rate of DC module

In the proposed architecture, we employ the dilated convolution (DC) module. Table 1 shows the impact of changing the dilated rate. From the experimental results it can be seen that the multiple dilated rates is better than the single dilated rate. We use the $\{1, 2, 4, 8\}$ and $\{1, 3, 5, 7\}$ dilated rate settings to achieve good performance.

3.2. The different upsampling methods

To explore the effect of the upsampling method on the network, we have added some ablation experiments, as the table 2. We experimentally found that the upsampling of deconvolution degrades the performance of the network. From the experimental results it can be seen that the bilinear upsampling method can achieve best performance. Therefore, we adopted the bilinear upsampling method, which effectively achieves better performance.

3.3. The effect of input size on the network

We have experimentally explore the effect of different input sizes on the network, as shown in Table 3. This is due to the fact that our network uses dilated convolution to help the network obtain a large receptive field, so the larger the size of the input the better our performance. However, the larger the input size, the larger the computer cost. Therefore, we make a compromise between performance and input size and choose an image size of 416×416 as the input size used for our final training.

Table 2. The effect of different upsampling methods on the network performance with an input of 96×96.

Upsampling Method	PSNR	SSIM
Deconvolution	26.56	0.7566
Transpose convolution	26.39	0.7471
Pixel-Shuffle	26.51	0.7509
Bicubic	26.45	0.7484
Bilinear	**26.71**	**0.7632**

Table 3. The effect of different input size images on the network. The following result is obtained from our training of 200 epochs.

Crop Size	128×128	256×256	320×320	416×416
Flops	45.33G	181.31G	283.29G	478.76G
Multi-Adds	120.34G	203.26G	317.6G	536.74G
PSNR	27.58	27.92	28.19	**29.16**
SSIM	0.834	0.836	0.845	**0.858**

4. The extended visual comparisons

Here we test the performance of different image deblurring methods on synthetic datasets. The visual results are shown in Figures 2 and 3 on the GoPro datasets [7].

References

[1] SWATI BIRARE and DR NALBALWAR. Review on super resolution of images using wavelet transform. *International Journal of Engineering Science and Technology*, 2, 12 2010. 1

[2] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia. Scale-recurrent network for deep image deblurring. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8174–8182, 2018. 3, 4

[3] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8877–8886, 2019. 3, 4

[4] H. Gao, X. Tao, X. Shen, and J. Jia. Dynamic scene deblurring with parameter selective sharing and nested skip connections. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3843–3851, 2019. 3, 4

[5] Dongwon Park, Dong Un Kang, Jisoo Kim, and Se Young Chun. Multi-temporal recurrent neural networks for progressive non-uniform single image deblurring with incremental temporal training. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision – ECCV 2020*. Springer International Publishing, 2020. 3, 4

[6] Kaihao Zhang, Wenhan Luo, Yiran Zhong, Lin Ma, Bjorn Stenger, Wei Liu, and Hongdong Li. Deblurring by realistic blurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2737–2746, 2020. 3, 4

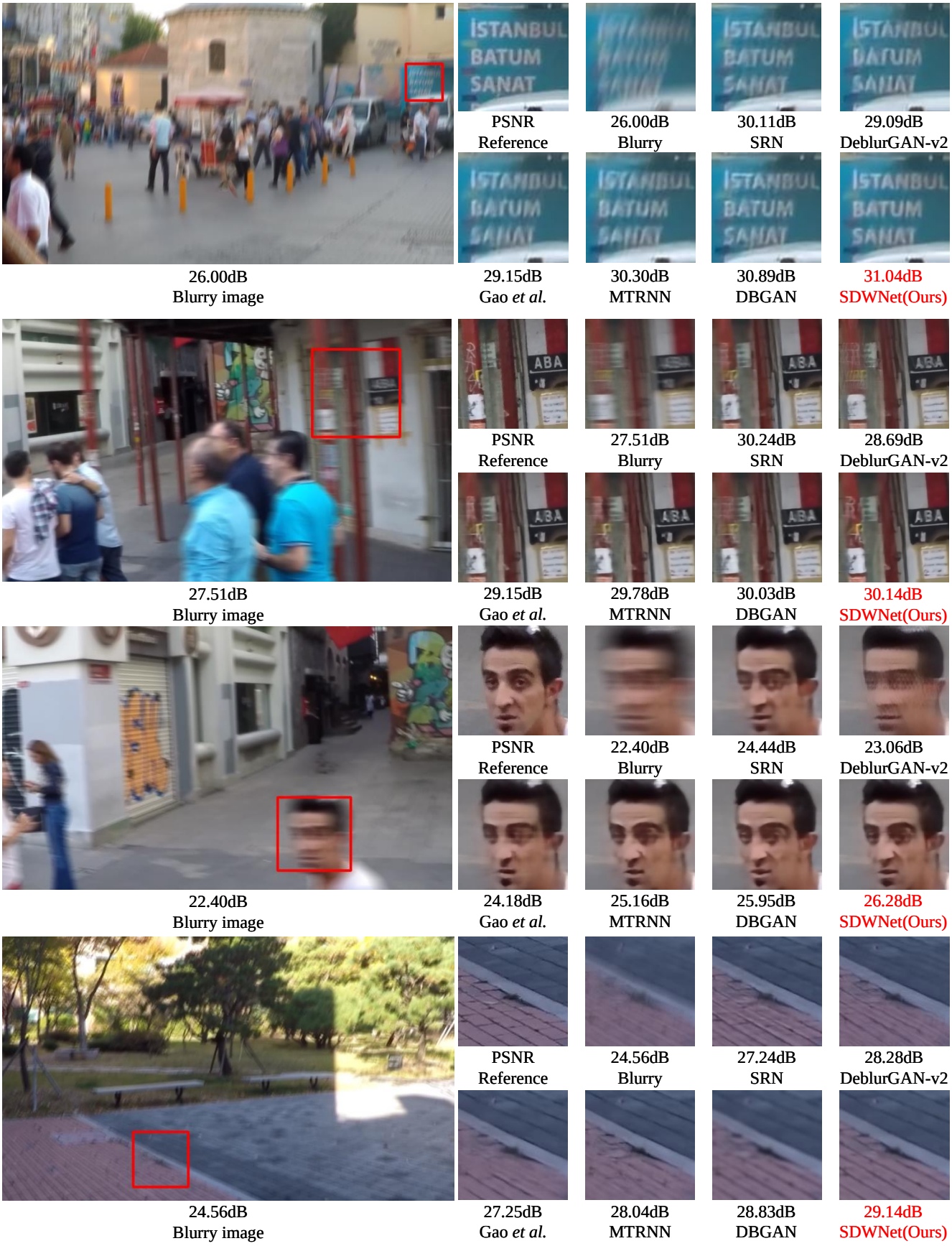

Figure 2. Qualitative comparison with the leading deblurring algorithms: SRN [2], DeblurGAN-v2 [3], Gao *et al.*[4], MTRNN[5], and DBGAN [6]. From the figure, we can see that our method can generate the right and clear details of the image.

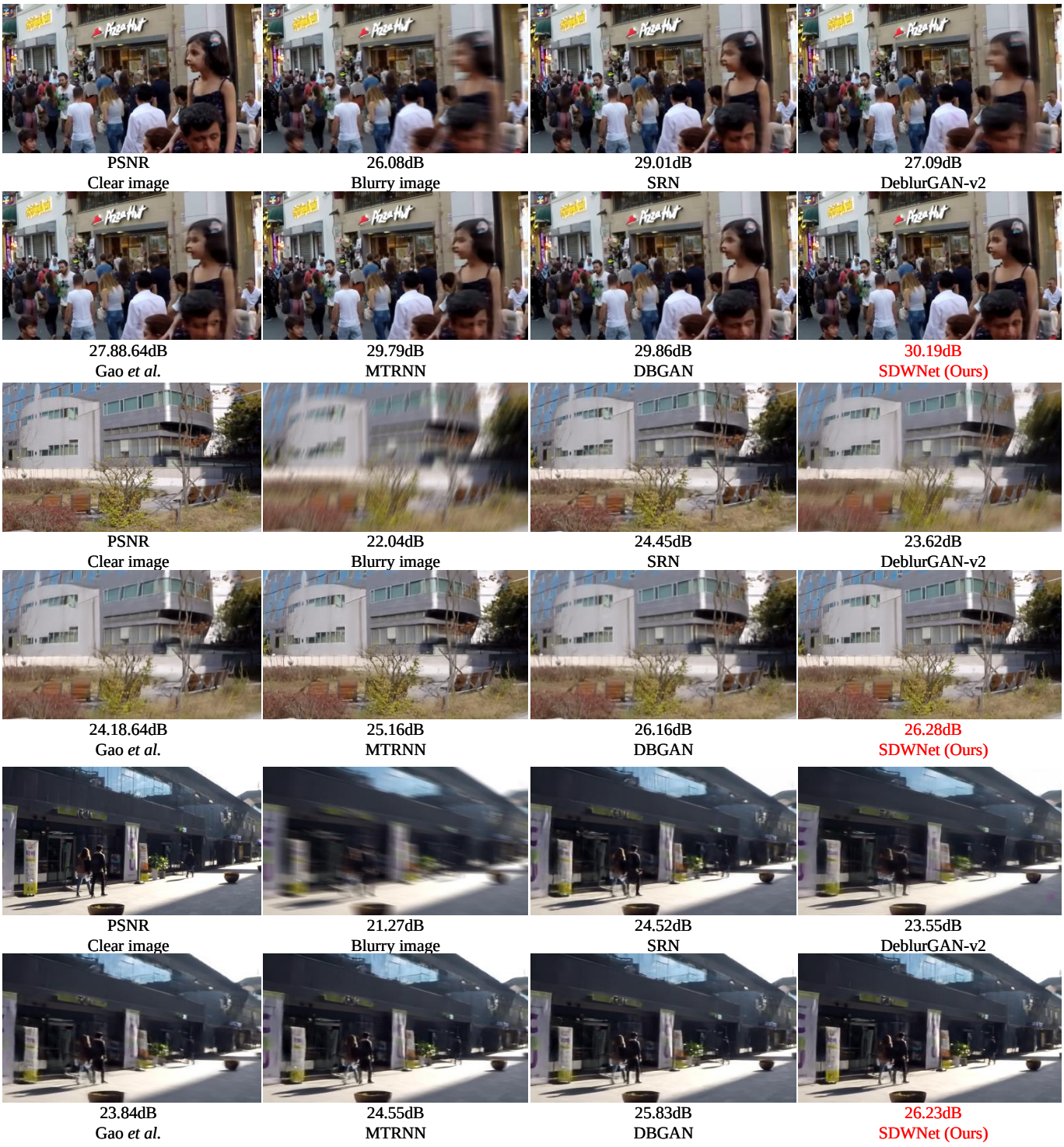

Figure 3. Visual qualitative comparison with the sate-of-the-art deblurring algorithms: SRN [2], DeblurGAN-v2 [3], Gao *et al.* [4], MTRNN[5], and DBGAN [6]. From the figure, we can see that our method can generate the right and clear details of the image.

[7] S. Nah, T. H. Kim, and K. M. Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 257–265, 2017. 2