

DPW-SDNet: Dual Pixel-Wavelet Domain Deep CNNs for Soft Decoding of JPEG-Compressed Images

Honggang Chen
Sichuan University
honggang_chen@yeah.net

Xiaohai He
Sichuan University
hxx@scu.edu.cn

Linbo Qing
Sichuan University
qing_lb@scu.edu.cn

Shuhua Xiong
Sichuan University
xiongsh@scu.edu.cn

Truong Q. Nguyen
UC San Diego
tqn001@eng.ucsd.edu

Abstract

JPEG is one of the widely used lossy compression methods. JPEG-compressed images usually suffer from compression artifacts including blocking and blurring, especially at low bit-rates. Soft decoding is an effective solution to improve the quality of compressed images without changing codec or introducing extra coding bits. Inspired by the excellent performance of the deep convolutional neural networks (CNNs) on both low-level and high-level computer vision problems, we develop a dual pixel-wavelet domain deep CNNs-based soft decoding network for JPEG-compressed images, namely DPW-SDNet. The pixel domain deep network takes the four downsampled versions of the compressed image to form a 4-channel input and outputs a pixel domain prediction, while the wavelet domain deep network uses the 1-level discrete wavelet transformation (DWT) coefficients to form a 4-channel input to produce a DWT domain prediction. The pixel domain and wavelet domain estimates are combined to generate the final soft decoded result. Experimental results demonstrate the superiority of the proposed DPW-SDNet over several state-of-the-art compression artifacts reduction algorithms.

1. Introduction

The number of devices with high-resolution camera increases significantly over the last few years, with the introduction of smart phones and IoT (Internet of Things) devices. Limited by the transmission bandwidth and storage capacity, these images and videos are compressed. As shown in Fig. 1, compressed images usually suffer from compression artifacts due to the information loss in the lossy compression process, especially at low bit-rates. In addition to poor perceptual quality, compression artifact-

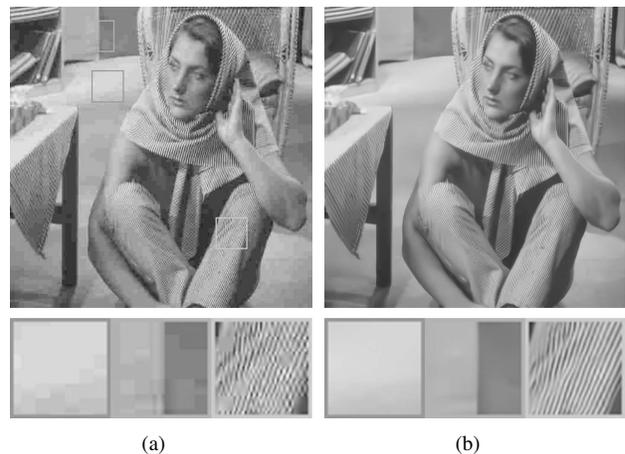


Figure 1. Illustrations of compression artifacts and soft decoding. (a) JPEG-compressed image in the case of QF = 10 (PSNR = 25.79 dB, SSIM = 0.7621, PSNR-B = 23.48 dB); (b) Soft decoded result of (a) using the developed DPW-SDNet (PSNR = 28.22 dB, SSIM = 0.8376, PSNR-B = 27.84 dB).

s also reduce the accuracy of other processing steps such as object detection and classification. Therefore, it is necessary to improve the quality of compressed images. This paper focuses on the soft decoding of JPEG images due to the fact that the JPEG is one of the commonly used compression standards for still images.

In recent years, many works investigate the restoration of JPEG images, aiming to remove compression artifacts and enhance the perceptual quality and objective assessment scores. In literature, the restoration procedure is usually referred to as soft decoding [21, 22], deblocking [20, 33], or compression artifacts reduction [5, 10]. In this paper, we use these terms interchangeably. Inspired by the excellent performance of the deep convolutional neural networks (CNNs) on various computer vision problems, we propose a

dual pixel-wavelet domain deep CNNs-based soft decoding network for JPEG-compressed images, namely DPW-SDNet. From Fig. 1 that illustrates a restored image by the proposed DPW-SDNet, we can observe that most of the compression artifacts are removed and some missing textures are recovered. Overall, the main contribution of this work is a dual-branch deep CNN that can reduce compression artifacts in both the pixel domain and wavelet domain. More specifically, our contributions are two folds:

- We develop an effective and efficient soft decoding method for JPEG-compressed images using dual pixel-wavelet domain deep CNNs. The combination of the pixel domain and wavelet domain predictions leads to better soft decoding performance.
- We reshape the compressed image and its 1-level discrete wavelet transformation (DWT) coefficients into two tensors with smaller size, which are used as the inputs to the pixel and wavelet sub-networks, respectively. By performing soft decoding on two smaller tensors, the DPW-SDNet achieves state-of-the-art performance while maintaining efficiency.

The rest of this paper is organized as follows. We describe the related work in the next section. The proposed soft decoding algorithm is presented in Section 3. Experiments are shown in Section 4. Finally, Section 5 concludes this paper.

2. Related Work

Let \mathbf{X} and \mathbf{Y} be the original uncompressed image and the corresponding JPEG-compressed version, respectively. Given the compressed image \mathbf{Y} , the goal of soft decoding is to produce an estimate that is as close as possible to the original image \mathbf{X} . Existing methods for soft decoding of JPEG-compressed images can be roughly split into three categories: enhancement-based, restoration-based, and learning-based methods.

The enhancement-based methods usually remove compression artifacts via performing pixel domain or transform domain filtering. For instance, Foi et al. [7] proposed a shape-adaptive discrete cosine transformation (DCT)-based image filtering, yielding excellent performance on deblocking and deringing of compressed images. Zhai et al. [31] proposed to reduce blocking artifacts via postfiltering in shifted windows of image blocks. In [30], the authors developed an efficient artifacts reduction algorithm through joint DCT domain and spatial domain processing. Yoo et al. [29] proposed an inter-block correlation-based blocking artifacts reduction framework, in which the artifacts in flat regions and edge regions were removed using different strategies.

Compression artifacts reduction is formulated as an ill-posed inverse problem for the restoration-based soft decoding methods, where the prior knowledge about high-quality

images, compression algorithms, and compression parameters is used to assist the restoration process [2, 4, 13, 20, 21, 22, 23, 24, 25, 32, 33, 36, 37, 38]. For instance, in [25], the original image and compression distortion were modeled as a high-order Markov random field and spatially correlated Gaussian noise, respectively. Non-local self-similarity property was widely used in deblocking algorithms. In general, the low-rank [20, 24, 33, 36] and group sparse representation [32, 38] were applied to model this property. In [2, 21, 22, 23, 32, 38], sparsity was utilized as an image prior to regularize the restored image. The graph model was used in the deblocking methods proposed in [13] and [21]. In some works [21, 22, 33, 36, 38], the quantization constraint on DCT coefficients was applied to restrain the resultant image. In particular, Dar et al. [4] designed a sequential denoising-based soft decoding algorithm, where the existing state-of-the-art denoising method was used to construct a regularization. On the whole, most of the restoration-based soft decoding methods are time-consuming to some extent due to the complex optimization process.

Recently, excellent results were obtained by deep learning-based approaches [1, 3, 5, 8, 9, 10, 19, 27, 34]. Dong et al. [5] developed a shallow CNN for compression artifacts reduction on the basis of the network for super-resolution [6]. The authors of [5] found that it is hard to train a network beyond four layers in low-level vision tasks. To address this issue, Kim et al. [17] introduced the residual learning technique and designed a very deep network of twenty layers for single image super-resolution. In [34], Zhang et al. presented a very deep network via incorporating the residual learning and batch normalization for a series of general image denoising problems, including denoising, super-resolution, and deblocking. Li et al. [19] combined the skip connection and residual learning to ease the network training process. Cavigelli et al. [1] developed a deep compression artifacts reduction network with a multi-scale loss function. In [3], Chen and Pock proposed a trainable nonlinear reaction diffusion model for efficient image restoration. Inspired by the success of the dual DCT-pixel domain sparse coding [22], the authors of [9] and [27] designed dual-domain networks for the deblocking of JPEG images. More recently, some works aim to improve the perceptual quality of compressed images [8, 10]. Overall, deep learning-based approaches show obvious superiority over conventional soft decoding methods in terms of both the restoration performance and running time¹.

Inspired by the success of the wavelet domain networks for super-resolution [11, 14], we present a dual pixel-wavelet domain deep CNN for the soft decoding of JPEG-compressed images in this paper. The proposed DPW-

¹ In general, the deep learning-based image restoration approaches are time-consuming in model training phase but efficient in testing phase. In this paper, the running time refers to the time cost in testing phase only.

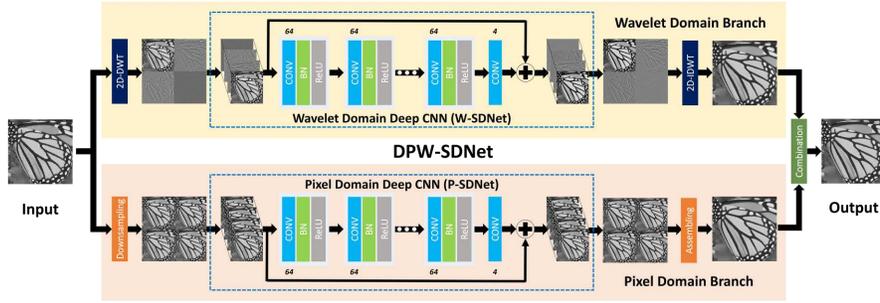


Figure 2. Flowchart of the proposed DPW-SDNet. The DPW-SDNet reduces compression artifacts in dual pixel-wavelet domain. The depths of the P-SDNet and W-SDNet are set to 20. The number next to each convolutional layer represents the number of kernels, and all of the convolutional layers in DPW-SDNet have the same kernel size of 3×3 .

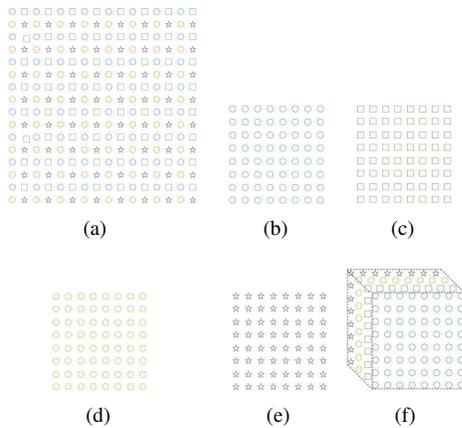


Figure 3. Illustration of the reversible downsampling process used in the pixel domain soft decoding branch. (a) The input image (size: $m \times n$, here $m = n = 16$); (b)-(e) Different downsampled versions of (a) (size: $\frac{m}{2} \times \frac{n}{2}$); (f) The tensor composed of (b)-(e) (size: $\frac{m}{2} \times \frac{n}{2} \times 4$). Note that this downsampling process is reversible.

SDNet is different from previous deep learning-based soft decoding algorithms in the following aspects: 1) The DPW-SDNet consists of two parallel branches that perform restoration in the pixel domain and wavelet domain, respectively. 2) The DPW-SDNet takes two tensors as the network inputs rather than the original compressed image and DWT coefficients. Experiments show that the DPW-SDNet achieves competitive restoration performance and execution speed on JPEG-compressed images. Moreover, the extensions of the proposed DPW-SDNet to other compression standards are straightforward.

3. Proposed DPW-SDNet

As outlined in Fig. 2, the proposed DPW-SDNet composes of two parallel branches: the pixel domain soft decoding branch and the wavelet domain soft decoding branch. The network in the pixel domain branch (namely P-SDNet)

removes compression artifacts in pixel domain directly, while the network in the wavelet domain branch (namely W-SDNet) performs restoration in wavelet domain. The pixel domain and wavelet domain estimates are combined to generate the final soft decoded result. Note that we do not directly use the original compressed image and its DWT sub-bands as the inputs of the two sub-networks. In the following sections, more details about the DPW-SDNet are presented. For convenience, we assume that the input \mathbf{Y} is a gray-scaled image of size $m \times n$ where m, n are both even.

3.1. The Pixel Domain Branch

In the pixel domain branch (shown in the bottom half of Fig. 2), first the compressed image \mathbf{Y} is downsampled to generate four downsampled sub-images of size $\frac{m}{2} \times \frac{n}{2}$. Since we have to recover an image that has the same size with the input, a reversible downsampling strategy is used in this process as [35]. Fig. 3 illustrates the reversible downsampling process. Given \mathbf{Y} , the pixels located at $(2i+1, 2j+1)$, $(2i+1, 2j+2)$, $(2i+2, 2j+1)$, and $(2i+2, 2j+2)$ ($i = 0, 1, 2, \dots, \frac{m}{2}-1$, $j = 0, 1, 2, \dots, \frac{n}{2}-1$) are respectively sampled to form four different sub-images, which are concatenated to constitute a tensor of size $\frac{m}{2} \times \frac{n}{2} \times 4$. Then, the tensor is fed into the pixel domain deep CNN. At least two benefits can be achieved by using a smaller tensor as the input of a deep CNN. First, a smaller input means lower computational complexity. In addition, working on the downsampled images can enlarge the receptive field, which is beneficial to restoration process.

For convenience, we name the pixel domain deep CNN P-SDNet. The input and output of the P-SDNet are tensors. The D -layer P-SDNet consists of two kinds of blocks. The first $(D-1)$ blocks are ‘‘CONV+BN+ReLU’’, and the last block only includes a convolutional layer. Note that the abbreviation ‘‘CONV’’ represents a convolutional layer, ‘‘BN’’ denotes the batch normalization [15], and ‘‘ReLU’’ represents the rectified linear unit [18]. The kernel number of each convolutional layer is set to 64 except the last layer

that outputs a 4-channel residual image. The kernel size of each convolutional layers is set to 3×3 . In each layer, the zero padding strategy is adopted to keep all feature maps having the same size. Since the input and output of the P-SDNet are very similar, we adopt the residual learning [12] for stable and fast training. Hence, the training loss function of the P-SDNet is defined as

$$L_P(\Theta_P) = \frac{1}{2N} \sum_{i=1}^N \|(f_p(\mathbf{y}_i^{pt}; \Theta_P) + \mathbf{y}_i^{pt}) - \mathbf{x}_i^{pt}\|^2 \quad (1)$$

where the Θ_P represents all parameters in P-SDNet, $f_p(\mathbf{y}_i^{pt}; \Theta_P)$ is the predicted residual component, and $\{(\mathbf{y}_i^{pt}, \mathbf{x}_i^{pt})\}_{i=1}^N$ denotes N compressed-clean tensor pairs in the pixel domain.

Finally, the four feature maps in the output of P-SDNet are assembled according to the inverse process of the down-sampling procedure to form the pixel domain estimate.

3.2. The Wavelet Domain Branch

The framework of the wavelet domain branch is similar to the pixel domain branch. Given a compressed image \mathbf{Y} , we first conduct the 1-level 2-dimensional discrete wavelet transformation (2D-DWT) and obtain its four wavelet sub-bands coefficients. The size of each sub-band is $\frac{m}{2} \times \frac{n}{2}$. Similarly, the four wavelet sub-bands are concatenated to constitute a tensor of size $\frac{m}{2} \times \frac{n}{2} \times 4$, which is used as the input of the wavelet domain deep CNN, namely W-SDNet. By concatenating four wavelet sub-bands, the information in different sub-bands can be fused while keeping the consistency among them. Moreover, the computational cost can be reduced.

The architecture of the W-SDNet is set to be the same as the P-SDNet, including the network depth, number of kernels, and kernel size. Therefore, we do not introduce the W-SDNet in details to avoid redundancy. The main difference between the two sub-networks is that the W-SDNet predicts wavelet coefficients residual while the P-SDNet predicts pixel intensity residual. Correspondingly, the training loss function of the W-SDNet is defined as

$$L_W(\Theta_W) = \frac{1}{2N} \sum_{i=1}^N \|(f_w(\mathbf{y}_i^{wt}; \Theta_W) + \mathbf{y}_i^{wt}) - \mathbf{x}_i^{wt}\|^2 \quad (2)$$

where the Θ_W represents all parameters in W-SDNet, $f_w(\mathbf{y}_i^{wt}; \Theta_W)$ is the predicted residual component, and $\{(\mathbf{y}_i^{wt}, \mathbf{x}_i^{wt})\}_{i=1}^N$ denotes N compressed-clean tensor pairs in the wavelet domain.

The four feature maps in the output of W-SDNet are the wavelet sub-bands of the soft decoded image. Therefore, the 2-dimensional inverse discrete wavelet transformation (2D-IDWT) is performed on these coefficients to produce the wavelet domain estimate.

3.3. The Combination of the Dual-Branch

As mentioned above, the pixel domain and wavelet domain branches both produce a soft decoded version of the input image. Since the two predictions are generated in different spaces, they have their respective characteristics. Hence, combining them should improve the restoration performance further. There are many ways to fuse the two intermediate results. For example, we can design a network with a 2-channel input and a 1-channel output to combine them. Considering the computational complexity, the two estimates derived from the dual-domain are simply equally weighted to generate the final output in this work.

4. Experiments

In this section, we first introduce some implementation details, followed by experimental results.

4.1. Implementation Details

Training Data: The publicly available imageset BSD-S500² is used to train the DPW-SDNet. We adopt the data augmentation (rotation and downsampling) to generate more training images. For the P-SDNet, we extract training sample pairs from original images and the corresponding compressed images. Correspondingly, the 2D-DWT coefficients of the original images and compressed images are used to generate training sample pairs for the W-SDNet. We generate $N = 523,968$ training sample pairs for each sub-network, and the size of each sample is set to $31 \times 31 \times 4$.

Training Parameters: We use the Caffe package [16] to implement the proposed network, and the depths of P-SDNet and W-SDNet are set to 20 ($D = 20$). The stochastic gradient descent algorithm is adopted to optimize our networks. The batch size, weight decay, momentum are set to 64, 0.0001, and 0.9, respectively. The initial learning rate is set to 0.1, and it decreases by a factor of 10 every 10 epochs. The maximum number of iterations is set to 300,000 for both the pixel domain and wavelet domain sub-networks.

4.2. Soft Decoding Performance Evaluation

The DPW-SDNet is compared with five state-of-the-art soft decoding algorithms for JPEG-compressed images, including two restoration-based approaches (i.e., CONCOLOR [33] and D2SD [22]) and three deep learning-based algorithms (i.e., ARCNN [5], TNRD [3], and DnCNN-3 [34]). Referring to [34], two benchmark imagesets Classic5 and LIVE1 are used as test datasets. For the color images in the LIVE1 dataset, only the luminance components are processed. The MATLAB JPEG encoder is used to generate JPEG-compressed images at different quality

²Available: <https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/resources.html>

Table 1. Average PSNR (dB)/SSIM/PSNR-B (dB) scores of different soft decoding algorithms on Classic5 and LIVE1. The best and the second-best scores are highlighted in red and blue, respectively.

QF		10	20	30	40
Classic5	JPEG	27.82/0.7595/25.21	30.12/0.8344/27.50	31.48/0.8666/28.94	32.43/0.8849/29.92
	CONCOLOR [33]	29.24/0.7963/29.14	31.38/0.8541/31.18	32.70/0.8809/32.50	33.60/0.8961/33.36
	D2SD [22]	29.21/0.7960/28.87	31.47/0.8551/31.15	32.79/0.8813/32.40	33.66/0.8962/33.20
	ARCNN [5]	29.05/0.7929/28.78	31.16/0.8517/30.60	32.52/0.8806/32.00	33.33/0.8953/32.81
	TNRD [3]	29.28/0.7992/29.04	31.47/0.8576/31.05	32.78/0.8837/32.24	-
	DnCNN-3 [34]	29.40/0.8026/29.13	31.63/0.8610/31.19	32.90/0.8860/32.36	33.77/0.9003/33.20
	DPW-SDNet	29.74/0.8124/29.37	31.95/0.8663/31.42	33.22/0.8903/32.51	34.07/0.9039/33.24
LIVE1	JPEG	27.77/0.7730/25.34	30.08/0.8512/27.57	31.41/0.8852/28.93	32.36/0.9041/29.96
	CONCOLOR [33]	28.87/0.8018/28.76	31.08/0.8681/30.90	32.42/0.8985/32.16	33.39/0.9157/33.07
	D2SD [22]	28.83/0.8023/28.54	31.08/0.8690/30.80	32.41/0.8987/32.10	33.37/0.9156/33.06
	ARCNN [5]	29.04/0.8076/28.77	31.31/0.8733/30.79	32.73/0.9043/32.22	33.63/0.9198/33.14
	TNRD [3]	29.14/0.8111/28.88	31.46/0.8769/31.04	32.84/0.9059/32.28	-
	DnCNN-3 [34]	29.19/0.8123/28.91	31.59/0.8802/31.08	32.99/0.9090/32.35	33.96/0.9247/33.29
	DPW-SDNet	29.53/0.8210/29.13	31.90/0.8854/31.27	33.31/0.9130/32.52	34.30/0.9282/33.44

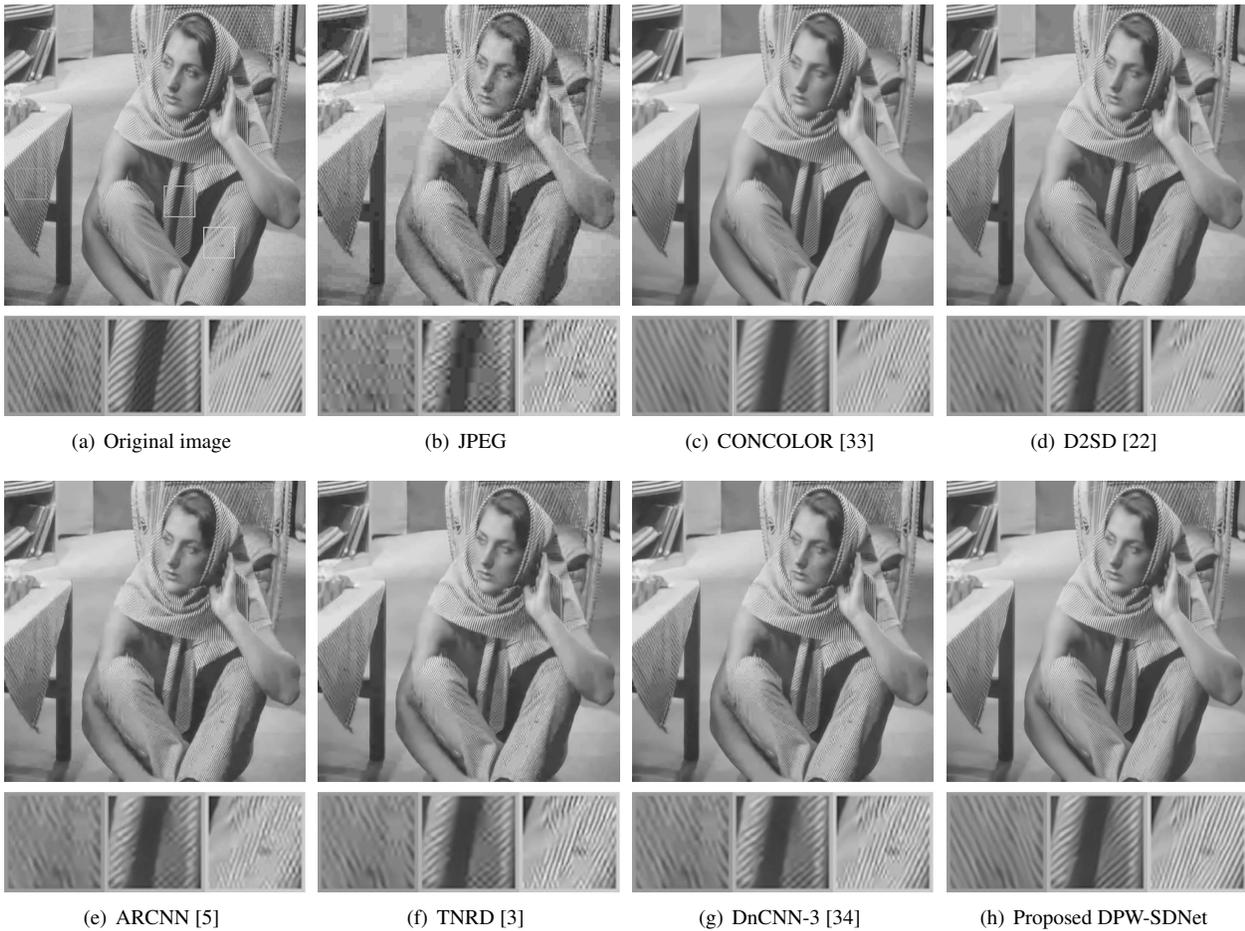


Figure 4. Visual quality comparison of different soft decoding methods on *Barbara* in the case of QF = 10. (a) Original image (PSNR (dB), SSIM, PSNR-B (dB)); (b) JPEG (25.79, 0.7621, 23.48); (c) CONCOLOR [33] (27.73, 0.8216, 27.63); (d) D2SD [22] (27.93, 0.8214, 27.64); (e) ARCNN [5] (26.92, 0.7967, 26.75); (f) TNRD [3] (27.24, 0.8099, 27.13); (g) DnCNN-3 [34] (27.58, 0.8161, 27.29); (h) Proposed DPW-SDNet (28.22, 0.8376, 27.84).

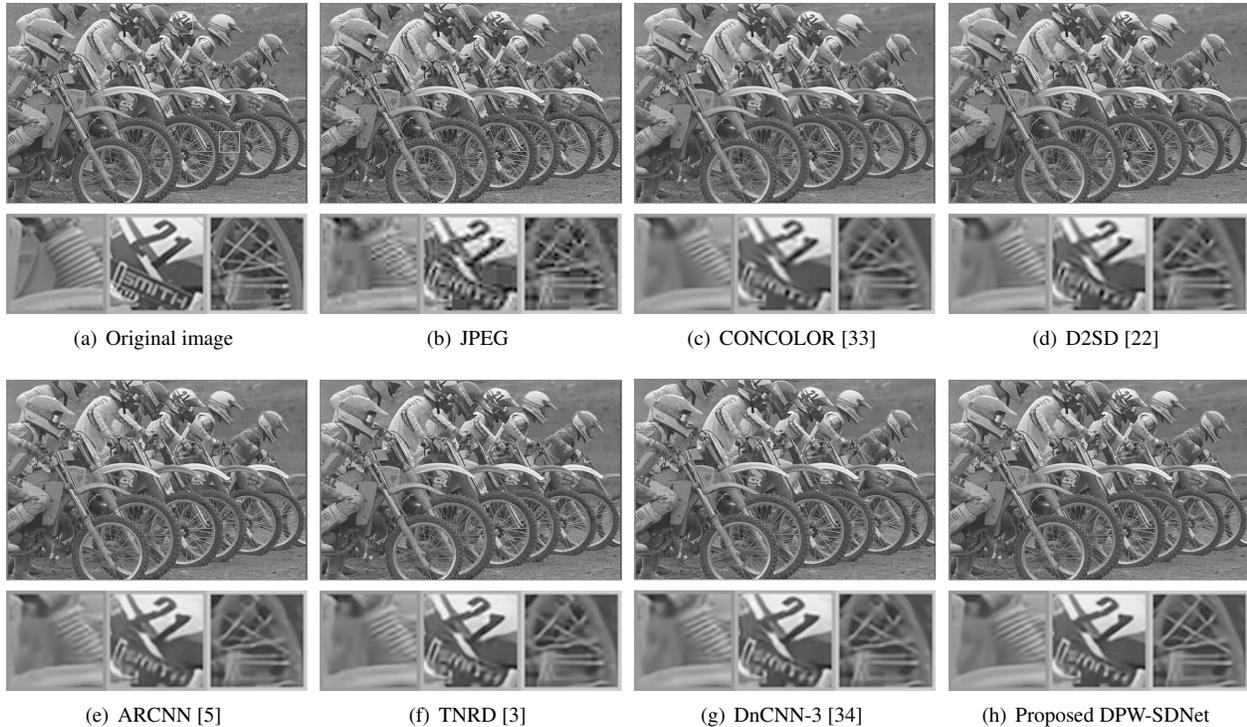


Figure 5. Visual quality comparison of different soft decoding methods on *Bike* in the case of $QF = 10$. (a) Original image (PSNR (dB), SSIM, PSNR-B (dB)); (b) JPEG (25.77, 0.7417, 23.02); (c) CONCOLOR [33] (27.00, 0.7801, 27.00); (d) D2SD [22] (27.11, 0.7859, 26.97); (e) ARCNN [5] (27.41, 0.7924, 27.11); (f) TNRD [3] (27.54, 0.7971, 27.22); (g) DnCNN-3 [34] (27.59, 0.7999, 27.28); (h) Proposed DPW-SDNet (28.04, 0.8133, 27.58).

factors (QFs). We compare the performance of these algorithms in the cases of $QF = 10, 20, 30,$ and 40 . For the DPW-SDNet, a dedicated model is trained for each QF. For the five competitors, we use the original codes and models provided by the authors.

Table 1 reports the objective assessment scores achieved by all tested algorithms, including the PSNR, SSIM [26], and PSNR-B [28]³. Note that the PSNR-B is a specifically developed assessment metric for blocky and deblocked images. It can be observed from Table 1 that the DPW-SDNet consistently outperforms the five competitors with considerable improvements. The only exception is the PSNR-B value on *Classic5* in the case of $QF = 40$, where the CONCOLOR [33] is superior to the DPW-SDNet. Overall, the DnCNN-3 [34] and TNRD [3] generate the second-best and the third-best results, respectively. The CONCOLOR [33], D2SD [22], and ARCNN [5] achieve comparable performance overall. On average, the proposed DPW-SDNet achieves about (0.30 ~ 0.34) dB PSNR gains, (0.0030 ~ 0.0098) SSIM gains, and (0.04 ~ 0.24) dB PSNR-B gains over the second-best approach DnCNN-3 [34]. The gains over the two restoration-based soft decod-

³ For the TNRD [3], the results at $QF = 40$ are not presented as the corresponding model is not available.

ing algorithms and ARCNN [5] are more significant. The improvements over state-of-the-art deblocking approaches demonstrate the effectiveness of the proposed DPW-SDNet.

One important aim of soft decoding algorithms is to recover images with high visual quality as JPEG-compressed images at high compression ratios usually suffer from severe artifacts. Therefore, some soft decoded images produced by different methods at $QF = 10$ are shown in Fig. 4, Fig. 5, and Fig. 6 in order to compare visual quality. It can be observed that most of the compression artifacts in JPEG images are removed by performing soft decoding on them. However, some soft decoded images are over-smoothed to some extent, or still suffer from visible artifacts. By contrast, the DPW-SDNet shows superiority in reducing artifacts and restoring details. The restored images using DPW-SDNet are more perceptually appealing, which can be seen from the highlighted regions. The results in this section verify that the DPW-SDNet not only achieves higher objective evaluation scores, but also produces better visual quality.

4.3. Discussion on Dual-Domain Soft Decoding

In DPW-SDNet, two parallel branches are used to restore the compressed image in the pixel domain and wavelet domain, respectively. It is meaningful to study the ability of



Figure 6. Visual quality comparison of different soft decoding methods on *Lighthouse3* in the case of $QF = 10$. (a) Original image (PSNR (dB), SSIM, PSNR-B (dB)); (b) JPEG (28.29, 0.7636, 25.98); (c) CONCOLOR [33] (29.77, 0.7976, 29.36); (d) D2SD [22] (29.77, 0.7977, 29.24); (e) ARCNN [5] (29.63, 0.7973, 29.19); (f) TNRD [3] (29.75, **0.8013**, 29.27); (g) DnCNN-3 [34] (**29.81**, 0.8007, **29.38**); (h) Proposed DPW-SDNet (**30.30**, **0.8104**, **29.76**).

Table 2. Average PSNR (dB)/SSIM/PSNR-B (dB) scores of different variants of the DPW-SDNet on Classic5 and LIVE1. The best scores are highlighted in **red**.

QF		10	20	30	40
Classic5	P-SDNet	29.69/0.8116/29.33	31.89/0.8657/31.39	33.18/0.8899/32.49	34.04/0.9036/33.22
	W-SDNet	29.70/0.8117/29.33	31.91/0.8660/31.37	33.18/0.8900/32.48	34.03/0.9036/33.21
	DPW-SDNet	29.74/0.8124/29.37	31.95/0.8663/31.42	33.22/0.8903/32.51	34.07/0.9039/33.24
LIVE1	P-SDNet	29.49/0.8203/29.10	31.86/0.8849/31.25	33.27/0.9126/32.49	34.26/0.9278/33.41
	W-SDNet	29.51/0.8205/29.11	31.87/0.8850/31.25	33.28/0.9127/32.50	34.26/0.9279/33.42
	DPW-SDNet	29.53/0.8210/29.13	31.90/0.8854/31.27	33.31/0.9130/32.52	34.30/0.9282/33.44

the two branches and discuss the effectiveness of the dual-domain combination. Table 2 presents the objective assessment scores of the DPW-SDNet and its two variants, i.e., the P-SDNet and W-SDNet. Here the P-SDNet represents that only the pixel domain branch is used to restore the compressed image, while the W-SDNet represents that only the wavelet domain branch is used.

It can be observed from Table 2 that both the P-SDNet and W-SDNet generate excellent restoration performance, which proves the ability of the presented network. Moreover, the gains of the DPW-SDNet over the P-SDNet and W-SDNet verify the effectiveness of the dual-domain soft decoding. Furthermore, it is believed that the fusion of the two branches could be more effective with a more complex combination method.

4.4. Discussion on Blind Soft Decoding

In above experiments, we use a dedicated model for each compression QF . To test the capacity of the DPW-SDNet further, we train a universal model for compressed images at different QFs . We refer to the universal model as the blind DPW-SDNet (B-DPW-SDNet), which is trained using the samples compressed at different QFs ⁴. In Section 4.2, DPW-SDNet and DnCNN-3 [34] perform the best and the second-best on the whole, respectively. Therefore, we compare the B-DPW-SDNet with them in Table 3.

As expected, the B-DPW-SDNet is slightly inferior to DPW-SDNet. However, in most cases, it still outperforms

⁴ Note that the same training dataset and the same number of training samples are used to train the universal model and the dedicated model.

Table 3. Comparisons of PSNR (dB)/SSIM/PSNR-B (dB) scores of the DnCNN-3 [34], DPW-SDNet, and B-DPW-SDNet on Classic5 and LIVE1. The best and the second-best scores are highlighted in red and blue, respectively.

QF		10	20	30	40
Classic5	DnCNN-3 [34]	29.40/0.8026/29.13	31.63/0.8610/31.19	32.90/0.8860/32.36	33.77/0.9003/33.20
	DPW-SDNet	29.74/0.8124/29.37	31.95/0.8663/31.42	33.22/0.8903/32.51	34.07/0.9039/33.24
	B-DPW-SDNet	29.69/0.8104/29.34	31.92/0.8660/31.39	33.18/0.8900/32.44	34.01/0.9035/33.19
LIVE1	DnCNN-3 [34]	29.19/0.8123/28.91	31.59/0.8802/31.08	32.99/0.9090/32.35	33.96/0.9247/33.29
	DPW-SDNet	29.53/0.8210/29.13	31.90/0.8854/31.27	33.31/0.9130/32.52	34.30/0.9282/33.44
	B-DPW-SDNet	29.48/0.8193/29.10	31.87/0.8849/31.26	33.27/0.9127/32.46	34.24/0.9278/33.38

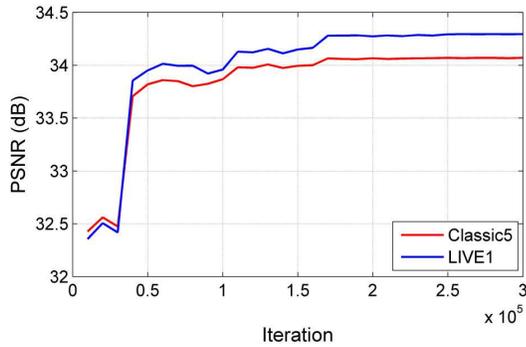


Figure 7. The PSNR (dB) values of DPW-SDNet on Classic5 and LIVE1 with different training iterations (QF = 40).

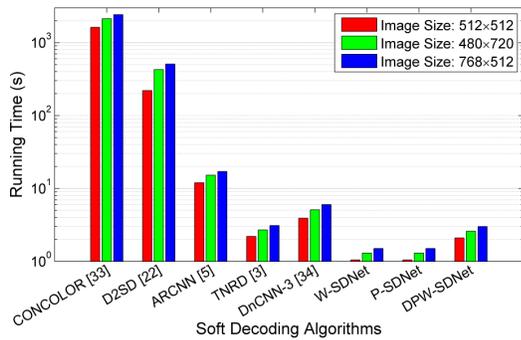


Figure 8. The running time (s) of different soft decoding algorithms on three representative image sizes in Classic5 and LIVE1.

DnCNN-3 [34] with obvious gains. Compared with DPW-SDNet, B-DPW-SDNet is more flexible and practical. Given QF, DPW-SDNet can be used to obtain better restoration performance, while B-DPW-SDNet can produce competitive results when the QF is unknown. Hence, one can select a proper model according to the practical application.

4.5. Empirical Study on Training Convergence and Running Time

In Fig. 7, we show the PSNR values of DPW-SDNet with different training iterations. The trends are similar for different QFs, so only the curves at QF = 40 are presented. It can be seen that the training converges after about 200,000

iterations. In our experiments, the maximum number of iterations is set to 300,000. The training of a single model takes about 9 hours on a GeForce GTX 1080 Ti GPU.

Running time is an important factor for a soft decoding algorithm. We run different deblocking methods on the same desktop computer with an Inter Core i7 CPU 4.2 GHz, 32GB RAM, and Matlab environment. Fig. 8 presents the execution time of different approaches on three representative image sizes in Classic5 and LIVE1⁵. It can be seen that the proposed P-SDNet and W-SDNet are the most efficient approaches. The DPW-SDNet costs about 2× time compared with the P-SDNet and W-SDNet, but it is still less time-consuming than other compared algorithms. Moreover, the execution speed of the DPW-SDNet can be greatly accelerated with a GPU.

5. Conclusion and Future Work

A dual pixel-wavelet domain deep network-based soft decoding framework is developed for JPEG-compressed images, namely DPW-SDNet. In DPW-SDNet, the compressed image is restored in both pixel and wavelet spaces using deep CNNs. In addition, we use 4-channel tensors as the inputs of our networks rather than the 2-dimensional images, which makes the DPW-SDNet efficient and effective. Experimental results on benchmark datasets demonstrate the effectiveness and efficiency of our soft decoding algorithm. Future work includes the extensions of the proposed DPW-SDNet to other image compression standards as well as other image restoration problems.

6. Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant 61471248, in part by the Fundamental Research Funds for the Central Universities under Grant 2012017yjsy159, and in part by the China Scholarship Council under Grant 201706240037. The authors thank Cheolhong An and Wenshu Zhan for helpful discussions.

⁵ In this experiment, the running time of the TNRD [3] is evaluated with the multi-threaded computation implementation.

References

- [1] L. Cavignoli, P. Hager, and L. Benini. CAS-CNN: A deep convolutional neural network for image compression artifact suppression. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, pages 752–759, 2017.
- [2] H. Chang, M. K. Ng, and T. Zeng. Reducing artifacts in jpeg decomposition via a learned dictionary. *IEEE Trans. Signal Process.*, 62(3):718–728, 2014.
- [3] Y. Chen and T. Pock. Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(6):1256–1272, 2017.
- [4] Y. Dar, A. M. Bruckstein, M. Elad, and R. Giryes. Post-processing of compressed images via sequential denoising. *IEEE Trans. Image Process.*, 25(7):3044–3058, 2016.
- [5] C. Dong, Y. Deng, C. L. Chen, and X. Tang. Compression artifacts reduction by a deep convolutional network. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 576–584, 2015.
- [6] C. Dong, C. C. Loy, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 184–199, 2014.
- [7] A. Foi, V. Katkovnik, and K. Egiazarian. Pointwise shape-adaptive dct for high-quality denoising and deblocking of grayscale and color images. *IEEE Trans. Image Process.*, 16(5):1395–1411, 2007.
- [8] L. Galteri, L. Seidenari, M. Bertini, and A. Del Bimbo. Deep generative adversarial compression artifact removal. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 4826–4835, 2017.
- [9] J. Guo and H. Chao. Building dual-domain representations for compression artifacts reduction. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 628–644, 2016.
- [10] J. Guo and H. Chao. One-to-many network for visually pleasing compression artifacts reduction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4867–4876, 2017.
- [11] T. Guo, H. S. Mousavi, T. H. Vu, and V. Monga. Deep wavelet prediction for image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1100–1109, 2017.
- [12] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.
- [13] W. Hu, G. Cheung, and M. Kazui. Graph-based dequantization of block-compressed piecewise smooth images. *IEEE Signal Process. Lett.*, 23(2):242–246, 2016.
- [14] H. Huang, R. He, Z. Sun, and T. Tan. Wavelet-SRNet: A wavelet-based cnn for multi-scale face super resolution. In *Proceedings of the IEEE Conference on Computer Vision (ICCV)*, pages 1689–1697, 2017.
- [15] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 448–456, 2015.
- [16] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 675–678, 2014.
- [17] J. Kim, J. Kwon Lee, and K. Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1646–1654, 2016.
- [18] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of the Neural Information Processing Systems Conference (NIPS)*, pages 1097–1105, 2012.
- [19] K. Li, B. Bare, and B. Yan. An efficient deep convolutional neural networks model for compressed image deblocking. In *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, pages 1320–1325, 2017.
- [20] T. Li, X. He, L. Qing, Q. Teng, and H. Chen. An iterative framework of cascaded deblocking and super-resolution for compressed images. *IEEE Trans. Multimedia*, 2017.
- [21] X. Liu, G. Cheung, X. Wu, and D. Zhao. Random walk graph laplacian-based smoothness prior for soft decoding of jpeg images. *IEEE Trans. Image Process.*, 26(2):509–524, 2017.
- [22] X. Liu, X. Wu, J. Zhou, and D. Zhao. Data-driven soft decoding of compressed images in dual transform-pixel domain. *IEEE Trans. Image Process.*, 25(4):1649–1659, 2016.
- [23] J. Mu, X. Zhang, R. Xiong, S. Ma, and W. Gao. Adaptive multi-dimension sparsity based coefficient estimation for compression artifact reduction. In *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6, 2016.
- [24] J. Ren, J. Liu, M. Li, W. Bai, and Z. Guo. Image blocking artifacts reduction via patch clustering and low-rank minimization. In *Data Compression Conference (DCC)*, pages 516–516, 2013.
- [25] D. Sun and W. K. Cham. Postprocessing of low bit-rate block dct coded images based on a fields of experts prior. *IEEE Trans. Image Process.*, 16(11):2743–2751, 2007.
- [26] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4):600–612, 2004.
- [27] Z. Wang, D. Liu, S. Chang, Q. Ling, Y. Yang, and T. S. Huang. D³: Deep dual-domain based fast restoration of jpeg-compressed images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2764–2772, 2016.
- [28] C. Yim and A. C. Bovik. Quality assessment of deblocked images. *IEEE Trans. Image Process.*, 20(1):88–98, 2011.
- [29] S. B. Yoo, K. Choi, and J. B. Ra. Post-processing for blocking artifact reduction based on inter-block correlation. *IEEE Trans. Multimedia*, 16(6):1536–1548, 2014.
- [30] G. Zhai, W. Zhang, X. Yang, W. Lin, and Y. Xu. Efficient deblocking with coefficient regularization, shape-adaptive

- filtering, and quantization constraint. *IEEE Trans. Multimedia*, 10(5):735–745, 2008.
- [31] G. Zhai, W. Zhang, X. Yang, W. Lin, and Y. Xu. Efficient image deblocking based on postfiltering in shifted windows. *IEEE Trans. Circuits Syst. Video Technol.*, 18(1):122–126, 2008.
- [32] J. Zhang, S. Ma, Y. Zhang, and W. Gao. Image deblocking using group-based sparse representation and quantization constraint prior. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pages 306–310, 2015.
- [33] J. Zhang, R. Xiong, C. Zhao, Y. Zhang, S. Ma, and W. Gao. CONCOLOR: Constrained non-convex low-rank model for image deblocking. *IEEE Trans. Image Process.*, 25(3):1246–1259, 2016.
- [34] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Trans. Image Process.*, 26(7):3142–3155, 2017.
- [35] K. Zhang, W. Zuo, and L. Zhang. FFDNet: Toward a fast and flexible solution for cnn based image denoising. *arXiv preprint arXiv:1710.04026*, 2017.
- [36] X. Zhang, W. Lin, R. Xiong, X. Liu, S. Ma, and W. Gao. Low-rank decomposition based restoration of compressed images via adaptive noise estimation. *IEEE Trans. Image Process.*, 25(9):4158–4171, 2016.
- [37] X. Zhang, R. Xiong, X. Fan, S. Ma, and W. Gao. Compression artifact reduction by overlapped-block transform coefficient estimation with block similarity. *IEEE Trans. Image Process.*, 22(12):4613–4626, 2013.
- [38] C. Zhao, J. Zhang, S. Ma, X. Fan, Y. Zhang, and W. Gao. Reducing image compression artifacts by structural sparse representation and quantization constraint prior. *IEEE Trans. Circuits Syst. Video Technol.*, 27(10):2057–2071, 2017.