NL-CALIC Soft Decoding Using Strict Constrained Wide-Activated Recurrent Residual Network

Yi Niu, *Member, IEEE*, Chang Liu, Mingming Ma, Fu Li, Zhiwen Chen, *Member, IEEE*, and Guangming Shi, *Fellow, IEEE*

Abstract— In this work, we propose a normalized Tanh activate strategy and a lightweight wide-activate recurrent structure to solve three key challenges of the soft-decoding of near-lossless codes: 1. How to add an effective strict constrained peak absolute error (PAE) boundary to the network; 2. An end-to-end solution that is suitable for different quantization steps (compression ratios). 3. Simple structure that favors the GPU and FPGA implementation. To this end, we propose a Wide-activated Recurrent structure with a normalized Tanh activate strategy for Soft-Decoding (WRSD). Experiments demonstrate the effectiveness of the proposed WRSD technique that WRSD outperforms better than the state-of-the-art soft decoders with less than 5% number of parameters, and every computation node of WRSD requires less than 64KB storage for the parameters which can be easily cached by most of the current consumer-level GPUs. Source code is available at https://github.com/dota-109/WRSD

Index Terms— NL-CALIC, soft decoding, normalized Tanh activate strategy, lightweight wide-activated recurrent structure.

I. INTRODUCTION

IN APPLICATIONS like remote sensing, telemedicine, Internet of things (IoT), etc., the image encoder is implemented by highly integrated circuit chips. It is intractable to update the on-chip image compression techniques for higher image quality. Therefore, soft decoding techniques have been investigated to improve the compressed image quality at the decoder side while maintaining the encoder unchanged.

Soft-decoding, in fact, is an ill-pose inverse problem and the pioneer soft-decoding techniques are based on explicit image modeling and optimization. A lot of sophisticated image models, like auto-regressive model [1], sparsity [2], [3], random walk graph [4], etc., are proposed for the soft decoding of the commonly used image/video codecs. Recently, researchers have been turned to adopt the popular convolutional neural networks (CNNs) to build data-driven and end-to-end soft-decoders instead of hand-crafted image modeling [5]–[12].
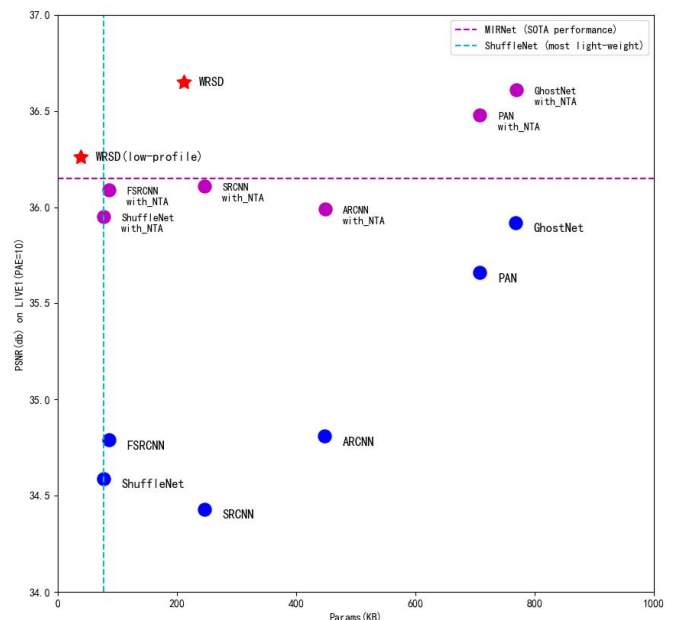
Fig. 1. Performance and Parameters comparison between WRSD and other lightweight models. "with_NTA" means the model is embedded in our NTA framework.

Due to the larger perception field and no need for explicit optimization, CNN-based soft-decoders outperform better than the traditional model-based soft-decoder in both image quality and executing time. Most CNN-based soft-decoding techniques regard the soft-decoding tasks as a common image restoration problem. This assumption works fine for the transform-based lossy codecs like JPEG [8], [9], [11], [12], JPEG 2000 [7]. However, as we are discussing now, this assumption DOES NOT suit the soft-decoding of the near-lossless codecs (NL-SD).

A. Background of Near-Lossless Coding (NLC)

Before the discussion about NL-SD, a background introduction of NLC is provided firstly to facilitate the subsequent descriptions and understanding of our new work. NLC is a special kind of lossy image compression technique. Besides pursuing low bit-rate and high visual quality like common lossy image codecs, the NLC techniques have an additional objective to constrain the peak absolute error (PAE) of every pixel. With the PAE constraint, NLC preserves the tiny weak local features which may be faded by the common lossy image

codecs. This makes NLC widely adopted in the scenarios like remote sensing, telemedicine, etc.. The name "near-lossless" comes from a straightforward NLC coding strategy that let the encoder send a quantized version of the input image losslessly, then lets the decoder adopt midpoint reconstruction to restore the lossless transferred quantized image. For example, let $\mathbf{I}$ and $\mathbf{S}$ denote the input image and the quantized image respectively. τ is the PAE constraint parameter that the quantization step $q = 2\tau + 1$ and $\mathbf{S} = \lfloor \mathbf{I}/q \rfloor$. The decoder restores $\hat{\mathbf{I}}$ via $\hat{\mathbf{I}} = \lfloor \mathbf{S} \times q + \tau \rfloor$, thus the PAE of $\hat{\mathbf{I}}$ is τ.

The SOTA near-lossless codec is NL-CALIC [13] which jointly optimize the quantization and encoding process in a predictive coding framework. The discussion of NL-CALIC is beyond the scope of this paper, and readers can simply regard the NL-CALIC soft-decoding process as the above image de-quantify problem, or refer to [13] for more details of the NL-CALIC coding and decoding process.

B. Challenges of NLC Soft-Decoding

Comparing with the CNN based soft decoder of common lossy codecs (CL-SD), NL-SD has four challenges:

Challenge 1: How to Constrain the L_∞ Bound Strictly: The compression distortion is caused by quantization. In common lossy codecs, the quantization is launched on transform domains like DCT or wavelet coefficients. Thus the CL-SD tasks are always regarded as a non-constraint optimization process. However, similar to the adaptive de-quantization problem, the quantization of near-lossless codecs is conducted on the pixel domain directly which restricts the peak absolute error (PAE) of every individual pixel. In this way, from the view of optimization, the NLC soft-decoding is a L_2 optimization process with L_∞ constraints during the restoration process:

$$\hat{\mathbf{I}} = \arg\min_{\hat{\mathbf{I}}} \left\| \hat{\mathbf{I}} - \mathbf{I} \right\|_2 \tag{1}$$

$$\text{s.t.} \ \left\| \hat{\mathbf{I}} - \mathbf{I} \right\|_\infty \le \tau \tag{2}$$

Challenge 2: One Model Per Compression Ratio: Like every lossy codec, NLC codecs should be worked at different network conditions. The only parameter which controls the compression ratio of a NLC is the PAE constraint which implies that the quantization step is $2\tau + 1$. In other words, the constraint varies for different compression ratios. For real applications, it is intractable to train a set of networks for different quantization steps, therefore, this makes the one-model-per-ratio tasks really challenging.

Challenge 3: The Dilemma Between Convergence and Overfitting/Gradient Vanish: As shown in Fig. 2, the NLC coding residual follows uniform distribution and the common lossy coding residual follows Gaussian/Laplacian distribution. In other words, the NL-CALIC soft decoder has to estimate the real value from the uniform distributed candidates, which is more challenging than estimating the real value from Gaussian/Laplacian distributed candidates.

The challenge can be explained from the view of distribution mapping [14]. Let us consider the soft decoding network as a distribution mapping function whose input follows uniform
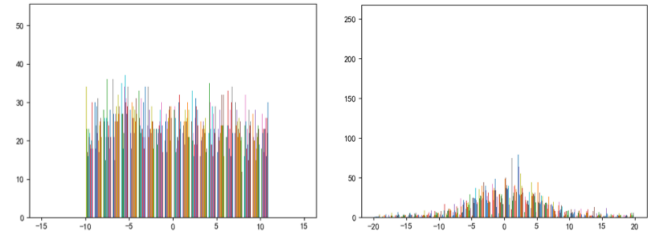
(a) The residual distribution of NLC coded image (NL-CALIC, τ=10)　(b) The residual distribution of common lossy coded image (JPEG, Q=10)

Fig. 2.　Comparison of coding residuals between NLC and common lossy codecs.

distribution and the output is the impulse distribution (the impulse located at real value), which is the distribution of the predicted image's residual. To easy for understanding, here we regard both the uniform distribution and the impulse distribution as two special Gaussian distributions whose variances tend to be infinity and zero respectively, and the convolution layer can be seen as a weighted average operation. In this way, the stacked convolution layers can be explained as decreasing the variance successively.

Generally speaking, regressing the variance from infinity to zero requires much more convolution layers than from a moderate value till convergence. The increase of network depth may bring overfitting and gradient vanish problems. Therefore, the dilemma between convergence, and overfitting/gradient vanish makes NLC soft decoding more challenging than the common image restoration tasks.

Challenge 4: Complexity and Executing Efficiency: Convergence, overfitting and gradient vanish are only challenges for the training process. In addition, in the implementation of the inference process, how to squeeze extra performance from the GPU devices to meet the system requirement of a large and sophisticated network is a bigger headache for the engineers like us.

C. Motivation and Contribution

The motivation of this work is to design a lightweight NLC soft decoding network whose structure favors the current consumer-level GPU devices and supports strict constraints and one-model-per-compression-ratio simultaneously.

Based on the above challenges and motivation, in this paper we propose a novel lightweight Wide-activated Recurrent structure with a normalized Tanh activate strategy for Soft-Decoding (WRSD) of NLC. The two main contributions of WRSD are:

- **The Normalized Tanh Activate (NTA) strategy:** For challenge 1 and challenge 2, we propose a normalized Tanh activate strategy which provides restrict L_∞ PAE constraints and support one-model-per-compression ratio simultaneously. The proposed NTA structure can be worked cooperatively with any image restoration net backbone to easily modify the general restoration net into NLC soft-decoder.
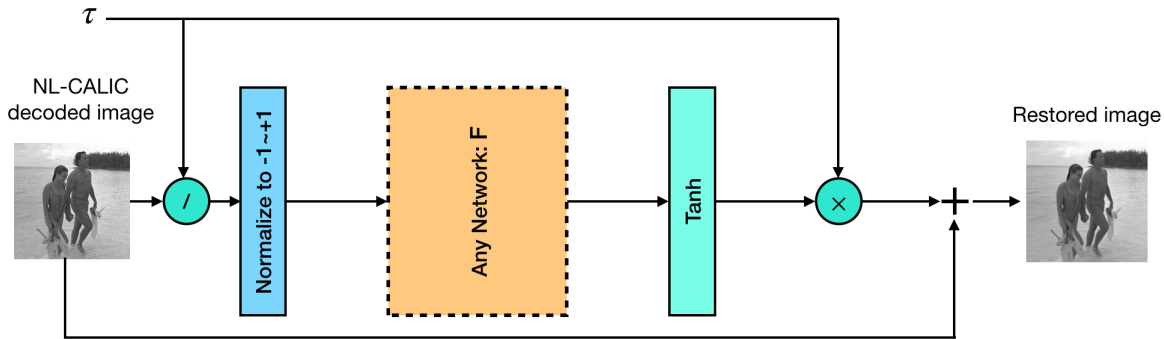
Fig. 3. The proposed NTA strategy, it consists of two operations about τ, normalization and Tanh activation function.

- **The lightweight wide-activate recurrent convolution structure:** We propose a lightweight wide-activated recurrent CNN model for NL-SD to solve challenge 3 and challenge 4. The recurrent and wide-activated structures not only speed up the convergence of the soft-decoding process while preventing the potential overfitting problem. In addition, all the computational nodes of the proposed WRSD technique require less than 64KB storage for the parameters which can be cached once-for-all by most of the current consumer-level GPU device's high-speed cache so-called shared memory.

Experimental results show that our model outperforms better than the state-of-the-art NL-SD method in both L_2 and L_∞ error. Our single network can decode multi-rate (2-10) coded images with faster speed, more restrict L_∞ constraint and fewer parameters than [15]. Moreover, our WRSD's weights number is even less than CL-SD method such as ARCNN [6]. By performing soft decoding with this unique structure, the WRSD achieves competitive performance while maintaining efficiency.

The rest of this paper is organized as follows. The related work is introduced in Section II. The method is proposed in Section III. Experiments are shown in Section IV. Finally, the conclusion is presented in Section V.

II. RELATED WORK

A. Traditional Soft Decoding Methods

In recent years, many image restoration methods have been proposed in the academic community to restore coded images [3], [4], [16]–[25]. These methods use transform coding or many kinds of image priors to obtain satisfactory restoration results. However, as introduced in section I, these kinds of soft decoders are not suitable for NL-SD that requires the L_∞ PAE constraint.

In the early years, there were several traditional soft decoders for the near-lossless codec. Zhou *et al.* [1] proposed a ℓ_2 based soft decoding method (named Soft CALIC) of ℓ_∞−decoded images, which shifted the task of improving the coding performance from the encoder to the decoder. In Soft CALIC, images are first compressed by NL-CALIC, and then a PAR-model-based soft decoding is applied on recovering images. This method greatly improved the low bit rate performance of near-lossless compression by regularizing

the quantization error. However, such an AR-based model is inferior to CNN-based models when solving data-driven image problems, the latter usually achieve better visual qualities (Fig. 15~Fig. 19) and objective qualities (Table. II) effects.

B. CNN Based Soft Decoding Methods

With the emergence of convolutional neural networks, deep learning is increasingly applied to image problems [26]–[29]. As far as we know, there are also several CNN-based soft decoding methods.

For CL-SD, Dong *et al.* [6] formulated a simple network for reducing compression artifacts; Wang *et al.* [30] introduced a Deep Dual-Domain model combined with both JPEG prior knowledge and sparse coding expertise, restoring JPEG-compressed images fast; He *et al.* [7] proposed a series of CNNs to performance JPEG-2000-compressed images at different coding bit-rates; Chen *et al.* [8] developed a dual pixel-wavelet domain deep CNNs-based soft decoding network for JPEG-compressed images; In [9], a deep wide-activated residual network was proposed to reduce blocking and color bleeding artifacts for colorful JPEG-compressed images; In [31], a JPEG compression artifacts reduction algorithm based on DCT coefficients prediction was proposed; In [32], Zini *et al.* proposed a deep residual auto encoder exploiting Residual-in-Residual Dense Blocks to remove artifacts in JPEG compressed images. In [33], a wavelet-supervision CNN with a large receptive field is proposed to solve the distortion of coded images.

For NL-SD, in 2019, a ℓ_∞−constrained CNN model was proposed for near-lossless compressed image restoration. Zhang *et al.* [15] used GAN to restore the near-lossless coded images by incorporating ℓ_∞−constrained loss in training loss. Regretfully, such a deep 16-layer and 4.5MB-parameter GAN model is only suitable for low bit rate (6,8,10) coded images rather than high bit rate (2,3,4) that is more important and complicated in the restoration of the near-lossless coded image. Besides, as a kind of weak constraint, ℓ_∞−constrained loss function is not sufficient to limit the restored interval of the coded image effectively.

C. Wide-Activated Identity Mapping

The proposal of Residual Network [34] makes the structure of the network deeper and deeper. In [35], a series of residual
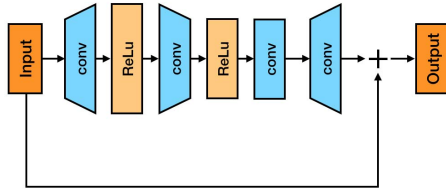
Fig. 4. Our residual block's structure.

blocks were studied and a 1000-layer-deep ResNet was proposed. In [36], Wu *et al.* analyzed the architecture of ResNet and designed a set of corresponding shallow networks with the results outperformed better than the previous deep residual networks. In [37], a wide residual network named WRN was proposed and achieved state-of-the-art on several datasets, indicating that the main advantage of residuals lies in residual blocks rather than depth. For instance, by adding more feature planes in each residual block, a wide 16-layer deep network actually has the same accuracy as a 1000-layer thin deep network. Nowadays, wide-activation based network was widely used in various image fields such as super resolution [38], deblocking [9] and restoration [10], [31]. The BPG/JPEG compressed images have been significantly restored in [10], [31] and the methods proposed in [9], [38] achieved state-of-the-art performance.

III. METHOD

In this section, we introduce our CNN-based image soft decoding network. The NTA strategy and the lightweight wide recurrent convolution structure will be shown firstly, then a total network's structure is given, finally, the loss function is introduced.

A. The Normalized Tanh Activate (NTA) Strategy

The main idea of NTA is inspired by the well-known ResNet. Instead of estimating the ground truth I directly from the NLC coded version $\hat{\mathbf{I}}$, NTA estimate the residuals $\mathbf{e}$ that:

$$\mathbf{I} = \hat{\mathbf{I}} + \mathbf{e} \tag{3}$$

Different from the common ResNet, the NLC decoded image has a restricted constraint that $\mathbf{e} \in [-\tau, \tau]$. So the NTA strategy estimates a normalized residual $\hat{\mathbf{e}}$ in $[-1, 1]$ that:

$$\mathbf{I} = \hat{\mathbf{I}} + \tau\hat{\mathbf{e}} \tag{4}$$

$$\mathbf{I}/\tau = \hat{\mathbf{I}}/\tau + \hat{\mathbf{e}} \tag{5}$$

If we consider the backbone network as a black-box function F, then we get the final NTA strategy as:

$$\mathbf{I}/\tau = \hat{\mathbf{I}}/\tau + Tanh(F(\hat{\mathbf{I}}/\tau)) \tag{6}$$

The normalization operator in Fig. 3 is a common strategy for numerical stability. The reader may get confused that the $/\tau$ operator before the normalization seems to be a waste operation because dividing a scalar before normalization is useless. We would like to maintain the $/\tau$ operator because it can be embedded into the original decoding process of

the NLC codec to avoid the mid-range dequantization of the DPCM residuals.

It can be observed that the Tanh activation is followed by $\times\tau$ operator ensures the restrict L_∞ bound of the restored image. In addition, since τ is the only compression ratio control parameter in NLC and the inference network is independent of τ, the one-model-per-compression-ratio can be achieved by training NTA with multi-ratio-compressed data sets.

B. The Lightweight Wide-Activate Recurrent Convolution Structure

Before the discussion of the technical details, we discuss our consideration of the term "lightweight".

In our understanding, "lightweight" means that the network should favor the current GPU structure with a small number of parameters, highly memory access efficiency, simple scheduling and fast executing time, which can be summarized as the following three aspects:

- The structure should be as simple as possible with minimized skip connections since every skip connection path requires large memory to buffer the intermediate feature map. (DenseNet is the most typical bad case)
- The structure should be as uniform as possible with minimized branches because branches will significantly increase the scheduling and synchronization time of the GPU. (inception and multi-scale attention are two typical bad examples)
- The number of parameters should be less than 64KB, or the computational task can be divided into several sub-tasks whose number of parameters is less than 64KB. This is because, in most of the current GPU devices, every stream process contains only 64KB high-speed cache so-called shared memory.

Now we consider challenge 3 which is the dilemma between convergence and overfitting. As we mentioned before, the NL-SD can be regarded as a distribution mapping problem that from infinity-variance Gaussian (uniform) distribution to zero-variance Gaussian distribution (impulsive). The regression of variance from infinity to zero requires multiple weighted average operators which implies a very deep neural network.

1) Why Recurrent CNN: Our main idea to simplify the potential very deep network is to use recurrent CNN structure which is borrowed from the DRCN [39]. Comparing with the common stacked CNN structure, the recurrent CNN structure has four advantages: 1) A significant reduction of the parameters which imply a strategy to reduce the model capacity for better generalization performances. 2) The reception field increases with every iteration which implies a multi-scale feature extraction. 3) A coarse-to-fine regression process which can be easily pruned to meet the different system requirements. 4) Reuse of the parameters can significantly reduce the data loading time since the parameters just need to be cached once.

However, the sophisticated recurrent CNN structure in DRCN is quite unfriendly to the GPU implementation because of the plenty of skip-connections and the multi-scale attention branches. This makes us investigate a more uniform recurrent
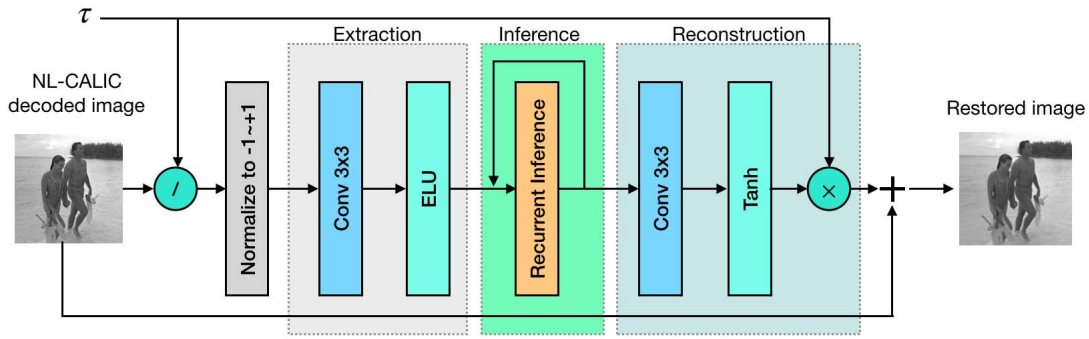
Fig. 5. Our network's structure. It consists of three parts: extraction part, recurrent inference part and reconstruction part. The unfold recurrent inference part is shown in Fig. 6.

TABLE I
COMPARISONS BETWEEN WRN AND WRSD

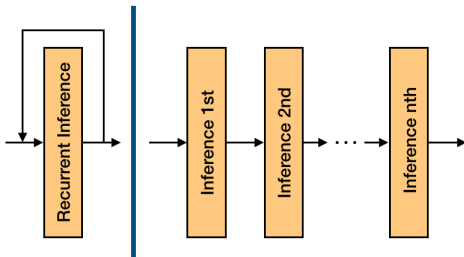| Method	Residual block's structure	Parameters		Average running time (s)
WRN	$3 \times 3 \times 64 \times 256$	576KB	1.125MB	0.428
	$3 \times 3 \times 256 \times 64$	576KB		
WRSD	$1 \times 1 \times 64 \times 256$	64KB	175KB	0.115
	$1 \times 1 \times 256 \times 48$	48KB		
	$3 \times 1 \times 48 \times 48$	63KB		
	$1 \times 3 \times 48 \times 64$			

Fig. 6. The recurrent inference part can be regarded as stack of several blocks, here we use the same filter to recurrent 12 times.

(a) DRCN-like[39, 41] recurrent structure.

(b) DR-ResNet-like[42] recurrent structure.

(c) The proposed WRSD's recurrent structure.

Fig. 7. Different recurrent-based structure. Obviously, the WRSD's recurrent structure is simple and efficient.

(a) PAE's comparison. (b) PSNR's comparison.

Fig. 8. Ablation study of NTA strategy under Kodak PhotoCD Dataset, WRSD_mse doesn't embed the NTA strategy.

(a) PSNR's comparison. (b) SSIM's comparison.

Fig. 9. Ablation study of two kind of residual blocks under Kodak PhotoCD Dataset. WRSD performs better in large PAE constraints.

structure to favor the GPU architecture and speed up the convergence of recurrent CNN.

2) Why Lightweight Wide-Activated: To accelerate the convergence, the recurrent CNN should contain sufficient weighting average operators which imply large convolution kernels. There are two ways to enlarge the convolution kernels: enlarge the spatial support or enlarge the channels. Theoretically, there are no differences between the two ways. However, to fully explore the high-speed cache of GPU, every convolution layer should be less than 64KB (the maximal kernel size is $3 \times 3 \times 42 \times 42$ for a single-precision floating-point).

Therefore, we choose to enlarge the channels of the convolution kernel instead of the spatial support. Guided by the two well-known concepts "expand & squeeze" and "separable convolution", we modify the wide-residual-network (WRN) [37] to a new lightweight wide-activated version which we call the Wide-Recurrent-Soft-Decoding (WRSD) network. The backbone network of WRSD is compared with the original WRN in Table. I. Both the two methods are tested under the Pytorch-GPU framework, and the GPU device is GTX 1050Ti.

To be specific, as shown in Fig. 4, the proposed WRSD backbone network has three sub-layers, and the memory cost of every sub-layer is less than 64KB. The first two layers form a simple wide-activated structur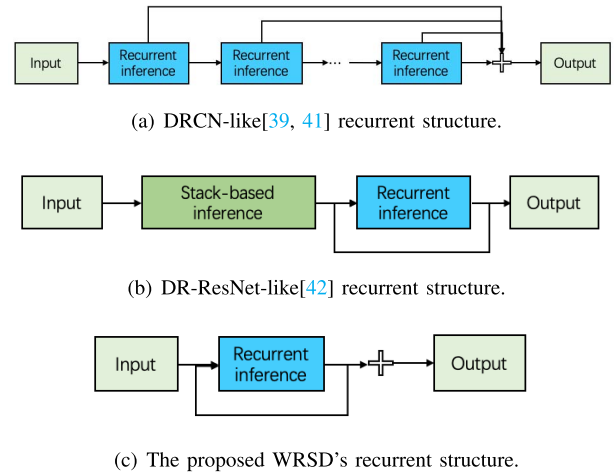e that uses 1×1 convolution kernels to expand and squeeze the feature map in the channel domain, respectively. The third sub-layer adopts a separable

TABLE II
AVERAGE PSNR (dB) SCORE/PAE ERROR OF DIFFERENT SOFT DECODING ALGORITHMS ON SEVERAL IMAGE DATASETS.
THE BEST AND THE SECOND SCORES ARE HIGHLIGHTED IN RED AND BLUE, RESPECTIVELY

	τ	2	3	4	5	6	7	8	9	10
Set5	Soft CALIC[1]	45.98 / 4.0	43.16 / 6.0	41.59 / 8.0	40.02 / 10.0	38.80 / 12.0	37.74 / 14.0	36.71 / 16.0	35.83 / 18.0	-/-
	ARCNN[6]	45.66 / 11.2	42.97 / 24.0	41.19 / 16.0	39.96 / 22.6	38.74 / 34.2	37.61 / 26.6	36.90 / 27.7	36.10 / 33.8	35.33 / 29.6
	SRCNN[40]	44.18 / 10.9	43.03 / 10.8	41.81 / 11.4	40.64 / 11.78	39.54 / 13.4	38.33 / 13.7	37.25 / 14.9	36.34 / 14.8	35.45 / 16.0
	FSRCNN[45]	43.34 / 23.3	42.49 / 24.8	41.54 / 24.2	40.44 / 25.3	39.40 / 25.3	38.36 / 23.1	37.31 / 28.2	36.34 / 24.1	35.45 / 24.7
	PAN[46]	46.68 / 7.4	44.51 / 7.7	42.87 / 8.1	41.52 / 10.0	40.45 / 11.3	39.49 / 13.1	38.72 / 14.2	37.88 / 16.4	37.20 / 17.2
	ShuffleNet[47]	45.26 / 8.1	43.64 / 9.1	42.06 / 9.3	40.72 / 10.4	39.60 / 11.8	36.45 / 13.8	37.52 / 15.2	36.67 / 16.2	35.87 / 17.6
	GhostNet[48]	46.66 / 7.3	44.53 / 7.4	42.90 / 8.4	41.56 / 10.4	40.49 / 11.2	39.60 / 12.8	38.82 / 14.6	37.98 / 15.7	37.31 / 17.5
	MIRNet[49]	46.93 / 5.5	44.70 / 6.5	43.06 / 7.8	41.72 / 10.2	40.65 / 11.0	39.79 / 12.7	39.05 / 14.4	38.25 / 15.8	37.59 / 17.1
	WRSD-L	46.90 / 3.4	44.39 / 5.1	42.61 / 6.8	41.17 / 8.5	40.01 / 10.2	39.17 / 11.9	38.31 / 13.5	37.48 / 15.1	36.76 / 16.9
	WRSD	47.12 / 3.4	44.66 / 5.1	42.96 / 6.8	41.58 / 8.5	40.47 / 10.2	39.67 / 11.9	38.86 / 13.6	38.04 / 15.3	37.38 / 17.0
Set14	Soft CALIC[1]	45.36 / 4.0	42.39 / 6.0	40.97 / 8.0	39.36 / 10.0	38.09 / 12.0	37.22 / 14.0	36.31 / 16.0	35.54 / 18.0	-/-
	ARCNN[6]	45.28 / 41.3	42.49 / 35.8	40.72 / 32.7	39.44 / 35.3	38.15 / 42.9	36.95 / 50.4	36.18 / 42.9	35.44 / 50.9	34.78 / 43.6
	SRCNN[40]	43.52 / 20.0	42.37 / 20.0	41.19 / 18.2	40.00 / 23.0	38.85 / 23.0	37.63 / 19.0	36.55 / 16.7	35.60 / 17.2	34.69 / 18.0
	FSRCNN[45]	42.35 / 27.1	41.55 / 25.7	40.67 / 27.8	39.70 / 28.7	38.65 / 25.6	37.71 / 29.6	36.67 / 29.5	35.74 / 28.5	34.86 / 29.3
	PAN[46]	45.94 / 9.4	43.66 / 9.0	41.97 / 10.0	40.63 / 10.8	39.51 / 11.8	38.00 / 13.6	37.19 / 15.1	36.43 / 17.2	35.76 / 18.2
	ShuffleNet[47]	44.64 / 11.1	42.96 / 11.2	41.40 / 12.1	40.04 / 14.8	38.85 / 14.6	37.52 / 15.2	36.49 / 16.6	35.59 / 17.1	34.76 / 18.7
	GhostNet[48]	45.92 / 8.9	43.65 / 9.4	41.99 / 10.5	40.66 / 11.4	39.54 / 11.7	38.20 / 13.2	37.38 / 15.2	36.62 / 17.1	35.96 / 18.3
	MIRNet[49]	46.33 / 5.4	43.87 / 6.5	42.17 / 8.0	40.83 / 9.7	39.73 / 11.5	38.39 / 13.5	37.61 / 15.6	36.86 / 16.8	36.17 / 18.4
	WRSD-L	46.25 / 3.4	43.62 / 5.1	41.82 / 6.8	40.38 / 8.5	39.18 / 10.2	38.25 / 11.9	37.36 / 13.6	36.63 / 15.3	35.97 / 17.0
	WRSD	46.38 / 3.4	43.81 / 5.1	42.07 / 6.8	40.69 / 8.5	39.55 / 10.2	38.56 / 11.9	37.72 / 13.6	37.00 / 15.3	36.40 / 17.0
BSD100	Soft CALIC[1]	45.21 / 4.0	42.06 / 6.0	40.60 / 8.0	38.86 / 10.0	37.53 / 12.0	36.63 / 14.0	35.56 / 16.0	34.86 / 18.0	-/-
	ARCNN[6]	45.19 / 47.2	42.38 / 43.2	40.49 / 45.2	39.23 / 44.0	37.98 / 49.2	36.81 / 70.3	36.08 / 48.7	35.26 / 55.9	34.60 / 59.7
	SRCNN[40]	43.73 / 18.5	42.41 / 19.0	41.04 / 20.0	39.71 / 22.2	38.43 / 22.8	37.25 / 17.1	36.14 / 21.6	35.13 / 21.2	34.18 / 25.0
	FSRCNN[45]	42.38 / 31.4	41.60 / 34.8	40.64 / 33.0	39.61 / 32.8	38.54 / 32.2	37.49 / 33.7	36.46 / 34.5	35.51 / 36.4	34.59 / 28.0
	PAN[46]	46.13 / 7.9	43.50 / 8.4	41.52 / 9.6	39.97 / 11.0	38.72 / 12.0	37.66 / 14.2	36.73 / 14.6	35.91 / 16.3	35.13 / 17.8
	ShuffleNet[47]	44.81 / 9.8	43.00 / 13.7	41.31 / 12.0	39.82 / 12.1	38.50 / 13.7	37.32 / 16.0	36.23 / 15.2	35.25 / 15.7	34.32 / 18.1
	GhostNet[48]	46.07 / 6.7	43.52 / 9.5	41.60 / 8.6	40.11 / 10.0	38.89 / 12.0	37.87 / 14.3	36.97 / 14.5	36.18 / 17.2	35.41 / 18.0
	MIRNet[49]	46.24 / 4.2	43.54 / 5.9	41.59 / 8.1	40.09 / 10.5	38.89 / 11.2	37.90 / 12.1	37.09 / 13.6	36.31 / 15.7	35.52 / 16.2
	WRSD-L	46.31 / 3.4	43.70 / 5.1	41.82 / 6.8	40.35 / 8.5	39.14 / 10.2	38.12 / 11.9	37.24 / 13.6	36.47 / 15.3	35.75 / 17.0
	WRSD	46.38 / 3.4	43.81 / 5.1	41.97 / 6.8	40.54 / 8.5	39.36 / 10.2	38.37 / 11.9	37.52 / 13.6	36.76 / 15.3	36.07 / 17.0
Kodak PhotoCD	Soft CALIC[1]	45.49 / 4.0	42.63 / 6.0	41.13 / 8.0	39.56 / 10.0	38.25 / 12.0	37.30 / 14.0	36.36 / 16.0	35.62 / 18.0	-/-
	ARCNN[6]	45.47 / 38.2	42.70 / 41.1	40.96 / 46.1	39.71 / 37.7	38.44 / 49.2	37.18 / 83.2	36.46 / 39.8	35.64 / 42.5	34.96 / 64.1
	SRCNN[40]	44.08 / 24.0	42.80 / 23.0	41.48 / 23.2	40.27 / 23.3	39.07 / 24.2	37.65 / 24.0	36.55 / 25.0	35.51 / 28.0	34.51 / 26.0
	FSRCNN[45]	42.75 / 34.7	41.94 / 33.0	40.96 / 30.6	39.98 / 32.0	38.93 / 36.1	37.89 / 32.3	36.88 / 32.0	35.88 / 36.3	34.90 / 34.0
	PAN[46]	45.79 / 12.4	43.69 / 12.5	42.15 / 12.8	40.90 / 14.0	39.84 / 14.0	38.16 / 15.4	37.34 / 16.3	36.57 / 18.5	35.80 / 19.0
	ShuffleNet[47]	44.77 / 11.8	43.07 / 12.2	41.48 / 13.3	40.20 / 14.5	39.05 / 15.0	37.65 / 15.4	36.62 / 15.8	35.66 / 17.4	34.71 / 19.1
	GhostNet[48]	45.69 / 12.1	43.67 / 12.6	42.16 / 13.1	40.92 / 13.8	39.87 / 14.2	38.32 / 15.2	37.57 / 16.8	36.80 / 18.5	36.06 / 20.1
	MIRNet[49]	46.52 / 4.7	44.16 / 6.3	42.48 / 8.4	41.19 / 10.7	40.12 / 12.0	38.65 / 13.0	37.94 / 15.0	37.17 / 17.0	36.31 / 19.0
	WRSD-L	46.31 / 3.40	43.84 / 5.1	42.05 / 6.8	40.68 / 8.5	39.48 / 10.2	38.60 / 11.9	37.83 / 13.6	37.06 / 15.3	36.29 / 17.0
	WRSD	46.48 / 3.40	44.11 / 5.1	42.38 / 6.8	41.07 / 8.5	39.93 / 10.2	38.98 / 11.9	38.21 / 13.6	37.48 / 15.3	36.79 / 17.0

convolution pair to form a 3×3 spatial convolution to expand the reception field.

Compare with the simplest $3 \times 3 \times 42 \times 42$ convolution solution, every coefficient of the WRSD feature map involves 608 $(64 + 256 + 3 \times 48 + 3 \times 48)$ weighted average operation, which is almost the double of the simple convolution $(3 \times 3 \times 42 = 378)$. More importantly, it increases the feature map channels from 42 to 64 which may carry more information in every recurrent loop.

C. The Main Architecture of the WRSD

To show the efficiency of the NTA strategy, the lightweight wide-activate recurrent convolution structure is embedded, as shown in Fig. 5. The total model consists of the extraction part, inference part and reconstruction part, respectively. Similar to [40], the extraction part uses one 3×3 conv layer to extract n-dimensional features from coded images, where $n = 64$.

The inference part finishes the prediction of error **e** between **I** and **Î** by using a recurrent residual block structure. The

unfold part is shown in Fig. 6. This structure helps our model learn **e** better without the need for additional new weight parameters.

For the reconstruction part, we adopt a 3×3 conv layer to decrease the channel firstly. Due to the high similarity of input and output, we add the residual learning [34] for fast training and convergence.

D. Loss Function

The proposed WRSD is an end-to-end optimization by minimizing the loss function:

$$L_{SUM} = L_{MSE} + \frac{\lambda}{2} \|\mathbf{W}\|_2 \qquad (7)$$

where the L_{MSE} is Mean Squared Error (MSE) loss, $\|\mathbf{W}\|_2$ is the L_2 regularization of weights parameters to prevent overfitting, λ is hyper-parameter.

1) MSE Loss: Our MSE loss has the following form:

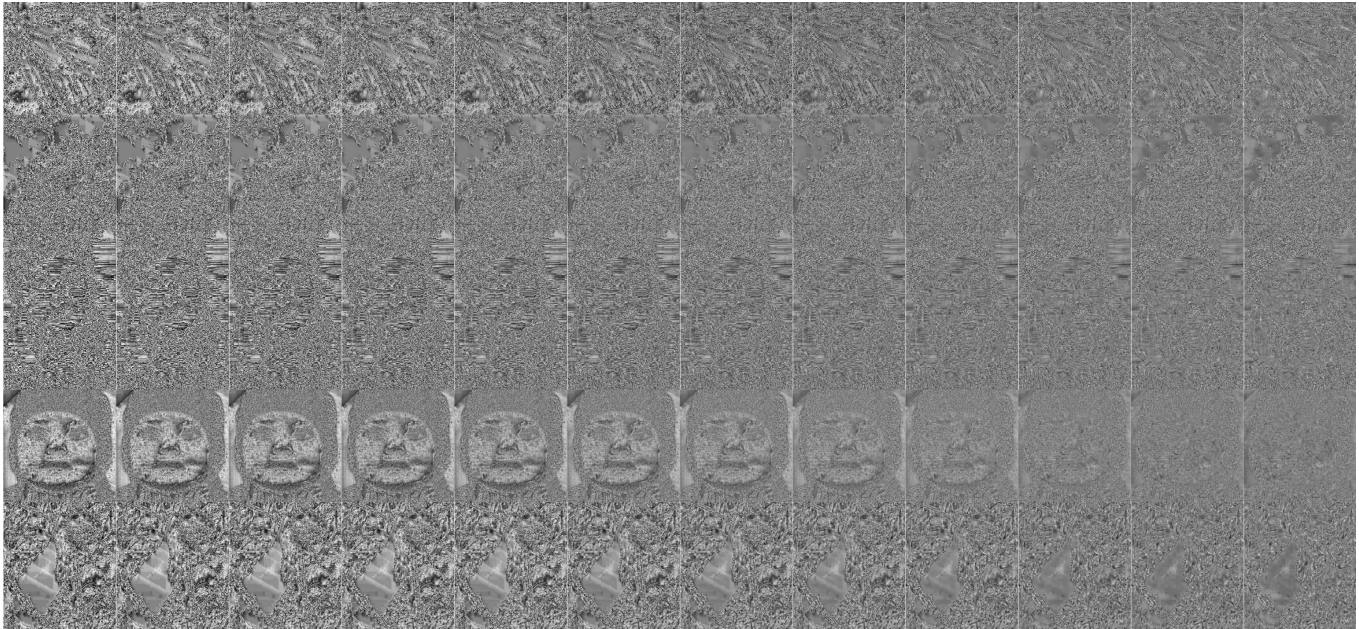$$L_{MSE} = \frac{1}{2 Rows \times Cols} \sum_{i=1}^{Rows \times Cols} \|x_i^r - x_i\|_2 \qquad (8)$$

Fig. 10.   Residual visualization of each recurrence in WRSD under Set5 dataset.

TABLE III

ABLATION STUDY OF NTA STRATEGY UNDER KODAK PHOTOCD DATASET, PSNR (DB) IS USED TO EVALUATE THESE METHODS. THE BEST SCORE IS HIGHLIGHTED IN **BOLD**

Method	Strategy	2	4	6	8	10
SRCNN	Original	44.08	41.48	39.07	36.55	34.51
	with NTA	**46.31**	**42.04**	**39.46**	**37.80**	**36.28**
FSRCNN	Original	42.75	40.96	38.93	36.88	34.90
	with NTA	**46.34**	**42.08**	**39.51**	**37.77**	**36.23**
ARCNN	Original	45.47	40.96	38.44	36.46	34.96
	with NTA	**46.28**	**41.96**	**39.32**	**37.67**	**36.17**
PAN	Original	45.79	42.15	39.84	37.34	35.80
	with NTA	**46.47**	**42.38**	**39.92**	**38.08**	**36.64**
ShuffleNet	Original	44.77	41.48	39.05	36.62	34.71
	with NTA	**46.26**	**41.91**	**39.25**	**37.61**	**36.07**
GhostNet	Original	45.69	42.16	39.87	37.57	36.06
	with NTA	**46.52**	**42.40**	**39.93**	**38.20**	**36.76**
MIRNet	Original	46.52	**42.48**	**40.12**	37.94	36.31
	with NTA	**46.56**	**42.48**	40.08	**38.43**	**37.03**

TABLE IV

COMPARISON OF AVERAGE PSNR (DB) AND MAXIMUM RESTORATION BOUNDARY BETWEEN OUR METHOD WRSD AND ℓ_∞−CNN [15]. THE BEST SCORE IS HIGHLIGHTED IN **BOLD**

DataSet	τ	Measure	MSE[15]	ℓ_∞−CNN[15]	WRSD
LIVE1 dataset	τ=6	PSNR	39.61	39.49	**39.84**
		SSIM	0.9731	**0.9745**	0.9737
		Bound	11.79	13.54	**10.20**
	τ=8	PSNR	37.84	37.72	**38.07**
		SSIM	0.9632	**0.9643**	0.9632
		Bound	15.37	18.34	**13.60**
	τ=10	PSNR	36.38	36.24	**36.65**
		SSIM	0.9518	**0.9532**	0.9520
		Bound	18.61	22.21	**17.00**
aerial image set	τ=6	PSNR	39.60	39.52	**39.97**
		SSIM	0.9620	0.9631	**0.9662**
		Bound	11.25	12.95	**10.20**
	τ=8	PSNR	37.84	37.76	**38.22**
		SSIM	0.9462	0.9475	**0.9519**
		Bound	14.89	17.24	**13.6**
	τ=10	PSNR	36.39	36.32	**36.82**
		SSIM	0.9293	0.9304	**0.9353**
		Bound	18.05	21.87	**16.99**

in which $x_i^r = F(y_i; w_i; b_i)$ is the pixel of restored image $\mathbf{I}^r$ by the WRSD model, w_i and b_i are part of **W**. MSE loss is widely used in a variety of image tasks and pursue the minimization of error energy while may not restrict error distribution. So we incorporate NTA strategy that has been described before to prove a high similarity in distribution.

E. Activation Functions

In our work, there have three kinds of activation functions, they are ReLu [43], ELU [44] and Tanh respectively. Firstly, the Tanh is adopted to constrain the error **e** that predicted by the CNN model; Secondly, due to the ReLu's gradient being easy to compute, it's used to guarantee the block's nonlinearity

and fast convergence during training; Finally, ELU can ensure that negative values are still activated, so it's used to extract features in the Extraction part.

F. Distinction Between WRSD and Other Recurrent Structure

Similar to DRCN [39], DRRN [41] and other CNN-based methods, we use recurrent structure to reduce params. For WRSD, there are still two differences with other recurrent-based CNN methods.

1) A Purely Recurrent Structure: Many recurrent-based CNN methods follow complicated structures that are shown in Fig. 7(a) and 7(b). DRCN [39] feeds each recurrence's output to the end; DR-RseNet [42] recurrently infers the stack block's
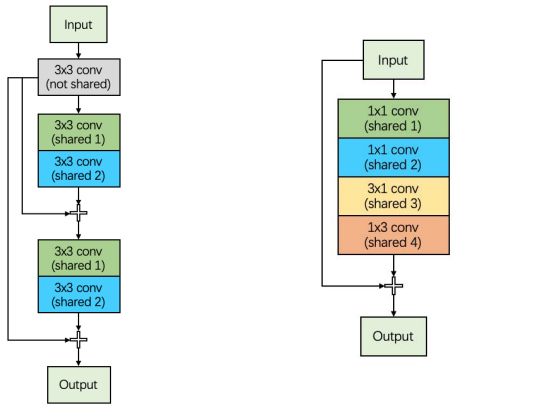
(a) DRRN's recursive block.   (b) The proposed WRSD's recurrent block.

Fig. 11.   Comparison between DRRN's and WRSD's recursive block. DRRN's conv kernels are not completely shared. For WRSD, all kernels in the block are totally shared and kernel sizes are small.
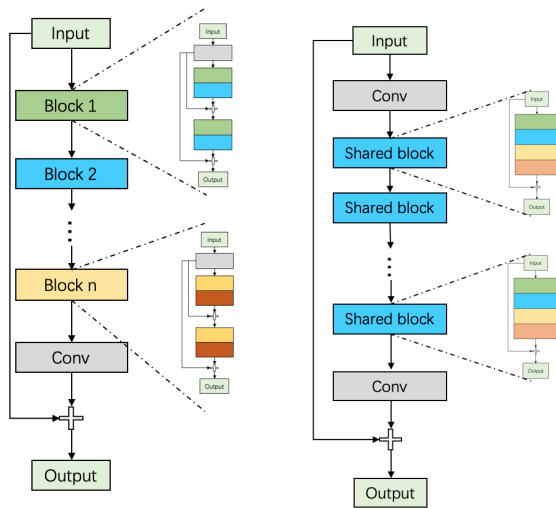
(a) DRRN's shared kernels only occur in the block.   (b) WRSD's each block has totally same params, which is a purely recurrent structure.

Fig. 12.   Comparison between DRRN and WRSD's recurrent way, network's structure have been simplified.

output. However, for WRSD, the architecture is consists of a purely recurrent structure, which has no redundant residual connections or stack structures.

2) Global Shared Weight Params: To improve performance, some conv kernels in the recurrent blocks may not be shared. For example, as shown in Fig. 11, the DRRN's recursive block consists of three kernels, the first of which is not shared. For WRSD, though there are four kernels in a recurrent block, most of them are small size kernels and all of them are totally shared.

Besides, DRRN's shared kernels only occur in the block. As shown in Fig. 12, though DRRN recursively restores the images, it is more like a stack-based CNN. However, for WRSD, the recurrent block's params are totally the same, which can be called a global totally shared recurrent CNN.
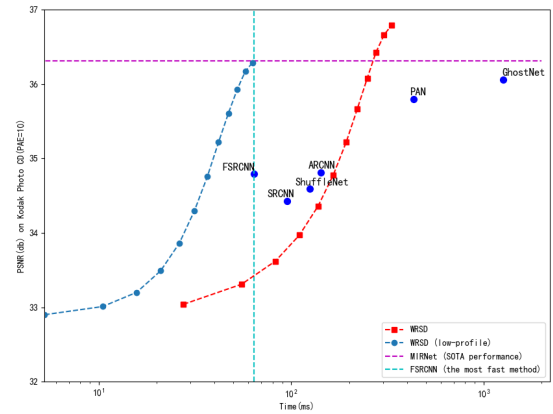

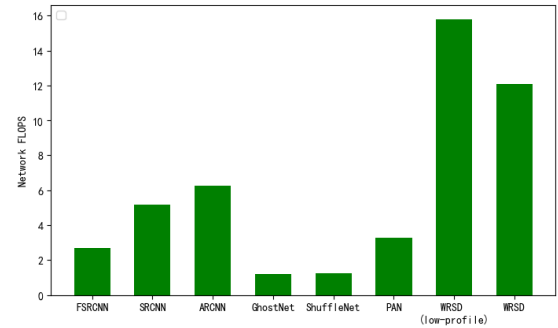Fig. 13.   Executing time comparison between WRSD and other methods.

Fig. 14.   Comparison of network FLOPS.

IV. EXPERIMENTS

Experimental results are presented to verify the effectiveness of our soft decoding method in this section. The WRSD is compared with some lightweight models, including ARCNN [6], SRCNN [40], FSRCNN [45], PAN [46], ShuffleNet [47] and GhostNet [48]. Soft CALIC [1] and MIRNet [49] are listed for reference. Soft CALIC is a traditional method; MIRNet is the baseline model. Noting that these CNN methods are trained on the same dataset as ours, and follow the same training way as WRSD.

A. Dataset

The database DIV2k is used for training our soft decoding network, it has 900 images about resolution $2K \times 1K$ in total. Here we set the batch size and patch size to 36 and 128, respectively. Kodak PhotoCD dataset, Set5, Set14, BSD100, LIVE1, Manga109, Urban100 and aerial image set are used as the test dataset. Set5, Set14, BSD100, Urban100 and Manga109 are popular datasets in the image super-resolution field, Kodak PhotoCD dataset has 24 images with resolution 768×512. To compare with $\ell_\infty -$CNN [15], we add LIVE1 and aerial image set as another test datasets. LIVE1 dataset has 29 images and most of them are the same as the Kodak PhotoCD dataset. The aerial image dataset has 180 remote sensing images with resolution 5000×5000. To keep consistency with others, we transform BSD100 and LIVE1's image format from ".jpg" and ".bmp" to ".png", respectively. The proposed WRSD is trained with single-channel images, but it

Fig. 15. Visual comparisons for PAE = 10.

can be easily applied to RGB images. NL-CALIC codec [13] is used to generate coded images at different quantization steps.

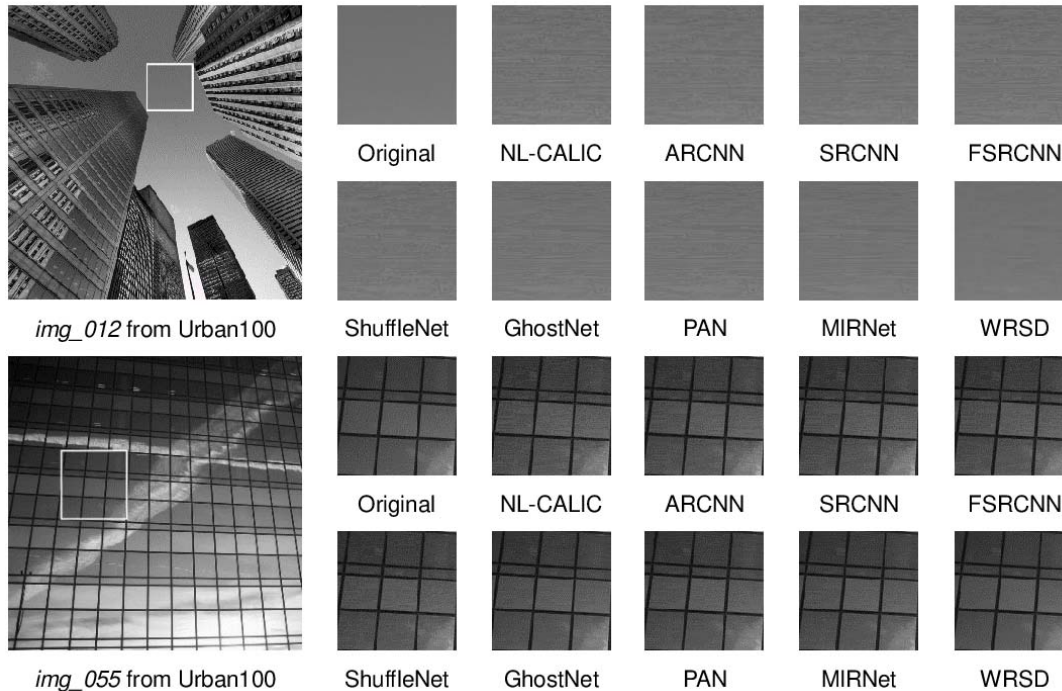We compare the performance of these algorithms in the case of $\tau = [2,10]$.

Fig. 16.   Visual comparisons for PAE = 10.

B. Training Details

Our WRSD's block loops 12 times, and most of the kernel sizes are 1×1, the number of parameters is only 46433, even less than ARCNN [6]. We randomly mix the training set with different quantization steps in each batch. Data augmentation is used for generating more training images. The learning rate is initialized to 5e-4, and it will be decreased if valid PSNR is not improved in 10 epochs. The minimum learning rate is 1e-5 and the learning rate decay is 0.5, the training epoch is set to 150. Training GPU device is RTX 2080Ti.

C. Ablation Study

1) The NTA Strategy: The ablation study of normalized Tanh function is shown in Fig. 8. WRSD means the originally proposed method, WRSD_mse means the WRSD without adopting NTA strategy. With the help of NTA strategy, the WRSD achieves better PSNR restoration and more strict PAE constraints.

2) Lightweight Wide-Activate Residual Block: To show the efficiency of the proposed lightweight wide-activate residual block, we compare it with the original residual block of WRN, both the above residual blocks' structures are shown in Table. I. From Fig. 9 we can know that WRSD has the same performance as WRN and even performs better in large PAE constraints while its parameters are much less than WRN's.

3) NTA With Lightweight CNNs and Baseline Model: For a more comprehensive ablation study, we also embed these CNN methods into the proposed NTA structure. For fairness of the compression, all these methods are constrained into $[-0.7\tau, 0.7\tau]$ as same as the proposed WRSD technique, thus all of the CNNs are with the same PAE constraints and we only show the PSNR performances. The performances are listed in

Table.III. It can be observed that taking advantage of the strict L_∞ constraint of NTA, most CNN+NTA techniques perform much better than the original CNNs. There is no doubt that the effectiveness of NTA has been proved.

4) WRSD-L, a Low-Profile Version: We propose a low-profile version of WRSD to favor low-end graphics whose shared memory is less than 48 KB, namely WRSD-L. The WRSD-L has only 39 KB by reducing the channels of WRSD from 64-256-48-48-64 to 16-64-12-12-16. The efficiency of WRSD will be verified from the comparisons of parameters, executing time and network FLOPS (NFLOPS). A detailed comparison please refer to the following sections.

D. Visualization of Recurrence

To further explain the decoding process of WRSD, we visualize the soft-decoding residuals after each recurrence. As shown in Fig. 10, as the increase of recurrence, the power of residual decreases and the residual pattern is turning to totally random noise gradually. This means that the soft-decoded images are turning to the original images after every recurrence. The PSNR results in Fig. 13 also prove this result.

E. Objective Quality Comparison

Table. II reports the PSNR assessment score achieved by all tested algorithms. It can be observed that the WRSD outperforms better than all the lightweight methods on most τ coefficients, and remains competitive performance when compared to MIRNet. These demonstrate the effectiveness of the proposed WRSD. Besides, the Urban100 and Manga109 datasets are also used to further evaluate the performances of WRSD. We compare WRSD with two SOTA lightweight

Fig. 17. Visual comparisons in coloful image (*kodim12* in Kodak PhotoCD dataset, $\tau = 10$). From top to bottom are the original image, NL-CALIC-coded image (PSNR/SSIM), GhostNet-decoded image (36.77/0.9772), PAN-decoded image (36.50/0.9758), MIRNet-decoded image (36.92/0.9780) and WRSD-decoded image (37.64/0.9816), respectively.

methods and MIRNet on $\tau = [2,10]$ and $[12,16]$, as shown in Table. V. PAN is a SOTA lightweight super-resolution method,

TABLE V
COMPARISON OF AVERAGE PSNR (DB) AND MAXIMUM RESTORATION BOUNDARY BETWEEN OUR METHOD WRSD AND PAN [46], GHOSTNET [48] AND MIRNET [49]. THE BEST SCORE IS HIGHLIGHTED IN **BOLD**

DataSet	τ	Measure	PAN	GhostNet	MIRNet	WRSD-L	WRSD
Urban100	2	PSNR	46.28	46.22	46.49	46.45	**46.59**
		Bound	10.3	9.7	4.4	**3.4**	**3.4**
	4	PSNR	41.92	42.00	42.11	42.07	**42.41**
		Bound	11.9	10.8	8.1	**6.8**	**6.8**
	6	PSNR	39.26	39.45	39.66	39.44	**39.98**
		Bound	13.1	13.1	12.5	**10.2**	**10.2**
	8	PSNR	37.38	37.62	38.06	37.54	**38.23**
		Bound	15.7	16.4	16.1	**13.6**	**13.6**
	10	PSNR	35.82	36.09	36.47	36.02	**36.84**
		Bound	19.1	20.1	18.6	**17.0**	**17.0**
	12	PSNR	34.42	34.67	34.85	34.77	**35.68**
		Bound	21.8	22.4	22.0	**20.4**	**20.4**
	14	PSNR	32.85	33.05	33.05	33.63	**34.64**
		Bound	24.4	23.9	24.4	**23.8**	**23.8**
	16	PSNR	31.41	31.55	31.48	32.71	**33.75**
		Bound	27.3	**25.7**	27.2	27.2	27.2
Manga109	2	PSNR	47.16	47.07	47.28	**47.44**	47.42
		Bound	8.8	9.3	4.7	**3.4**	**3.4**
	4	PSNR	42.71	42.92	43.06	43.22	**43.30**
		Bound	11.1	10.9	8.3	**6.8**	**6.8**
	6	PSNR	40.15	40.47	40.73	40.74	**40.99**
		Bound	12.7	12.6	11.7	**10.2**	**10.2**
	8	PSNR	38.37	38.72	39.18	38.92	**39.30**
		Bound	15.4	16.1	14.9	**13.6**	**13.6**
	10	PSNR	36.82	37.19	37.53	37.46	**37.97**
		Bound	18.4	19.1	18.0	**17.0**	**17.0**
	12	PSNR	35.29	35.62	35.55	36.18	**36.81**
		Bound	21.1	21.1	**19.7**	20.4	20.4
	14	PSNR	33.64	33.87	33.46	35.07	**35.77**
		Bound	23.1	23.1	**21.9**	23.8	23.8
	16	PSNR	32.06	32.19	31.68	34.08	**34.80**
		Bound	26.2	**24.5**	25.3	27.2	27.2

TABLE VI
CNN METHODS' PARAMETERS COMPARISON (KB)

WRSD	WRSD-L	ShuffleNet 76.8	FSRCNN 85.7	SRCNN 246.8	ARCNN 448.4
212	**39**	PAN 708.3	GhostNet 767.9	ℓ_∞−CNN 5120	MIRNet 130764

GhostNet performs competitive accuracy in NL-SD, MIRNet is our baseline model. Obviously, WRSD outperforms better than these three SOTA methods. Besides, WRSD also performs well in coefficients [12,16] that have not been trained.

Moreover, we compare WRSD with ℓ_∞−CNN [15] which is shown in Table. IV. Same as [15], we use LIVE1 and a set of aerial images as test images. In general, our model has better PSNR, lower PAE, and competitive SSIM.

F. Visual Quality Comparison

Another important goal of the soft decoding algorithm is to restore images with high visual quality as NL-CALIC-compressed images usually have severe block artifacts at high

Fig. 18. Visual comparison in colorful image (*UchiNoNyansDiary_000* in Manga109, $\tau = 10$). From upper left to lower right are the original image, NL-CALIC-coded image (PSNR/SSIM), PAN-decoded image (36.57/0.9806), GhostNet-decoded image (36.78/0.9804), MIRNet-decoded image (36.95/0.9809) and WRSD-decoded image (38.33/0.9816).

PAE. Therefore, some soft decoded images with different methods at $\tau = 10$ are shown in Fig. 15 and Fig. 16 for visual quality evaluation. In general, images restored by WRSD are more visually pleasing than other methods. In addition, to prove the proposed WRSD still performs well in colorful images, we compare it with SOTA lightweight methods and baseline model: PAN, GhostNet, and MIRNet. Fig. 17, Fig. 18 and 19 prove that WRSD also performs well in colorful images.

G. CNN Parameters Comparison

Due to adopting the recurrent structure and 1×1 convolution kernel, our WRSD achieves state-of-the-art near-lossless coded images' restoration while taking up very little storage space. The comparison results of all parameters are shown in Fig. 1 and Table. VI. Obviously, WRSD achieves SOTA performance while parameters are competitive. For the low-profile version of WRSD, has only 39 KB, and is also better than most light-weight competitors. Besides, benefited from the NTA strategy, other methods also achieve at least 0.5 dB improvements, which verifies the efficiency of NTA framework.

H. Executing Time Comparison

Firstly, let us show the comparison results of executing time before and after acceleration. As shown in Tab. VII, after acceleration, both the two versions of WRSD largely improve the speed. Besides, we provide a comprehensive comparison of executing time, as shown in Fig. 13. Since the WRSD restores images recurrently (12 times), the executing time consumption is consists of 12 points. For the low-profile version of WRSD, it achieves the fastest speed with only a few dB losses. Besides, WRSD is more flexible because of its recurrent structure. One can choose recurrent times based on the trade-off of performance and executing time.
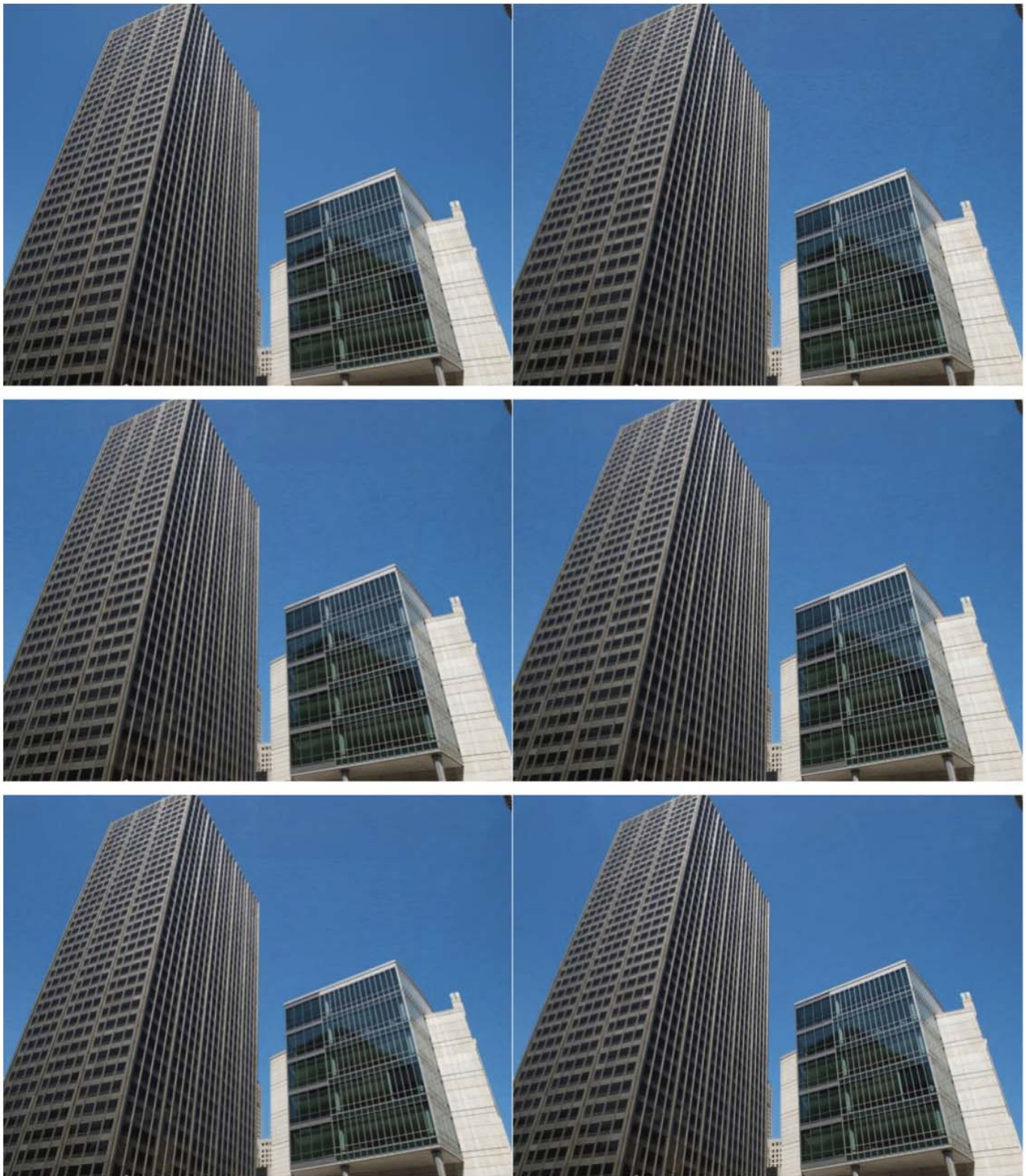
Fig. 19. Visual comparison in colorful image (*img_096* in Urban100, $\tau = 10$). From upper left to lower right are the original image, NL-CALIC-coded image (PSNR/SSIM), PAN-decoded image (35.59/0.9844), GhostNet-decoded image (36.01/0.9859), MIRNet-decoded image (36.57/0.9864) and WRSD-decoded image (38.12/0.9904).

I. Network FLOPS Comparison

Except for the comparison of parameters and executing time, in this section, we adopt a new metric to evaluate networks' performances, which is the network FLOPS (NFLOPS). FLOPS means the floating-point operations per second. Simply put, the larger the FLOPS, the better the method. The comparison of NFLOPS is shown in Fig. 14.

TABLE VII
COMPARISON OF EXECUTING TIME (MS) BEFORE
AND AFTER ACCELERATION

WRSD-L		WRSD	
Original	Accelerated	Original	Accelerated
179	65	1304	331

As we can see, compared with other light-weight methods, both the two versions of WRSD have much larger NFLOPS, which verifies the efficiency of WRSD.

V. CONCLUSION

In this paper, we propose a novel near-lossless soft-decoding network via normalized Tanh activate (NTA) strategy and a lightweight wide-activate recurrent convolution structure. The proposed WRSD techniques achieve the SOTA soft-decoding performances with a simple and GPU-favored structure. Firstly, the NTA strategy supports the strict L_∞ constraints and one-model-per-compression ratio simultaneously and can be worked cooperatively with any backbone network. This property made NTA strategy a generalized structure for all the L_∞ constrained scenarios like HDR imaging, image re-quantization, and image denoising of dark current noise. In addition, the lightweight wide-activate recurrent convolution structure, as an alternative structure for the stacked CNN structure, has the potential to be extended to the other low-level visual tasks like super-resolution, denoising. Due to its simple structure, efficient cache capability and flexibility time-complexity control, it is quite friendly for the SOC design.

However, it should be noticed that the current recurrent CNN structure seems only suitable for low-level visual tasks, we already did some attempts to extend the recurrent CNN structure to high-level visual tasks like image recognition, the performance is quite negative. This may be because high-level visual tasks need the extraction of semantic level features which implies large model capacity. The knowledge distillation and multi-level recurrent maybe two directions in our future work.

REFERENCES

[1] J. Zhou, X. Wu, and L. Zhang, "ℓ_2 Restoration of ℓ_∞-decoded images via soft-decision estimation," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4797–4807, Jun. 2012.

[2] Y. Li and J. Zhou, "Sparsity-driven reconstruction of ℓ_∞-decoded images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 4612–4616.

[3] X. Liu, G. Cheung, X. Wu, and D. Zhao, "Inter-block consistent soft decoding of JPEG images with sparsity and graph-signal smoothness priors," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 1628–1632.

[4] 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.*, vol. 26, no. 2, pp. 509–524, Feb. 2017.

[5] X. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in *Advances in Neural Information Processing Systems*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2016, pp. 2802–2810.

[6] C. Dong, Y. Deng, C. C. Loy, and X. Tang, "Compression artifacts reduction by a deep convolutional network," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 576–584.

[7] X. He, H. Chen, J. Chen, and L. Qing, "Soft decoding of JPEG 2000 compressed images using bit-rate-driven deep convolutional neural networks," in *Proc. IEEE Int. Conf. Inf. Automat. (ICIA)*, Jul. 2017, pp. 843–847.

[8] H. Chen, X. He, L. Qing, S. Xiong, and T. Q. Nguyen, "DPW-SDNet: Dual pixel-wavelet domain deep CNNs for soft decoding of JPEG-compressed images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 711–720.

[9] H. Chen, X. He, C. An, and T. Q. Nguyen, "Deep wide-activated residual network based joint blocking and color bleeding artifacts reduction for 4:2:0 JPEG-compressed images," *IEEE Signal Process. Lett.*, vol. 26, no. 1, pp. 79–83, Jan. 2019.

[10] Y. Fan, J. Yu, and T. S. Huang, "Wide-activated deep residual networks based restoration for bpg-compressed images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2018, pp. 2621–2624.

[11] P. Liu, H. Zhang, K. Zhang, L. Lin, and W. Zuo, "Multi-level wavelet-CNN for image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 773–782.

[12] L. Cavigelli, P. Hager, and L. Benini, "CAS-CNN: A deep convolutional neural network for image compression artifact suppression," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 752–759.

[13] X. Wu and N. Memon, "CALIC—A context based adaptive lossless image codec," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 4, May 1996, pp. 1890–1893.

[14] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: https://www.deeplearningbook.org

[15] X. Zhang and X. Wu, "Near-lossless ℓ_∞-constrained image decompression via deep neural network," in *Proc. Data Compress. Conf. (DCC)*, Mar. 2019, pp. 33–42.

[16] S. Liu and A. C. Bovik, "Efficient DCT-domain blind measurement and reduction of blocking artifacts," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 12, pp. 1139–1149, Dec. 2002.

[17] X. Liu, X. Wu, J. Zhou, and D. Zhao, "Data-driven sparsity-based restoration of jpeg-compressed images in dual transform-pixel domain," in *Computer Vision Pattern Recognition*. Boston, MA, USA: IEEE Computer Society, 2015.

[18] T. Komatsu, Y. Ueda, and T. Saito, "Super-resolution decoding of JPEG-compressed image data with the shrinkage in the redundant DCT domain," in *Proc. 28th Picture Coding Symp.*, Dec. 2010, pp. 114–117.

[19] Y. Niu, G.-M. Shi, X.-T. Wang, L.-Z. Wang, and D.-H. Gao, "JPEG stream soft-decoding technique based on autoregressive modeling," *J. China Universities Posts Telecommun.*, vol. 19, no. 5, pp. 115–123, Oct. 2012.

[20] S. I. Young, A. T. Naman, and D. Taubman, "COGL: Coefficient graph Laplacians for optimized jpeg image decoding," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 343–355, Jan. 2019.

[21] C. J. Lin, Y. T. Prasetyo, and R. Widyaningrum, "Eye movement measures for predicting eye gaze accuracy and symptoms in 2D and 3D displays," *Displays*, vol. 60, pp. 1–8, Dec. 2019.

[22] W. A. Hemmerich, A. Shahal, and H. Hecht, "Predictors of visually induced motion sickness in women," *Displays*, vol. 58, pp. 27–32, Jul. 2019.

[23] X. Ye *et al.*, "An optimization method for parameters configuration of the light field display based on subjective evaluation," *Displays*, vol. 62, Apr. 2020, Art. no. 101945.

[24] X. Liu, D. Zhai, D. Zhao, G. Zhai, and W. Gao, "Progressive image denoising through hybrid graph Laplacian regularization: A unified framework," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1491–1503, Apr. 2014.

[25] X. Liu, D. Zhao, R. Xiong, S. Ma, W. Gao, and H. Sun, "Image interpolation via regularized local linear regression," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3455–3469, Dec. 2011.

[26] X. Liu, D. Zhai, R. Chen, X. Ji, D. Zhao, and W. Gao, "Depth restoration from RGB-D data via joint adaptive regularization and thresholding on manifolds," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1068–1079, Mar. 2019.

[27] Y. Liu, S. Zhang, J. Xu, J. Yang, and Y.-W. Tai, "An accurate and lightweight method for human body image super-resolution," *IEEE Trans. Image Process.*, vol. 30, pp. 2888–2897, 2021.

[28] M. Ma, Y. Niu, and R. Li, "Bayesian network based general correspondence retrieval method for depth sensing with single-shot structured light," *Displays*, vol. 67, Apr. 2021, Art. no. 102001.

[29] X. Liu, G. Cheung, X. Ji, D. Zhao, and W. Gao, "Graph-based joint dequantization and contrast enhancement of poorly lit JPEG images," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1205–1219, Mar. 2019.

[30] 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 *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2764–2772.

[31] M. Sun, X. He, S. Xiong, C. Ren, and X. Li, "Reduction of JPEG compression artifacts based on DCT coefficients prediction," *Neurocomputing*, vol. 384, pp. 335–345, Apr. 2020.

[32] S. Zini, S. Bianco, and R. Schettini, "Deep residual autoencoder for blind universal JPEG restoration," *IEEE Access*, vol. 8, pp. 63283–63294, 2020.

[33] Z. Cao, T. Zhang, M. Liu, and H. Luo, "Wavelet-supervision convolutional neural network for restoration of JPEG-LS near lossless compression image," in *Proc. IEEE Asia Conf. Inf. Eng. (ACIE)*, Jan. 2021, pp. 32–36.

[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[35] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis.* Scottsdale, AZ, USA: Springer, 2016, pp. 630–645.

[36] Z. Wu, C. Shen, and A. Van Den Hengel, "Wider or deeper: Revisiting the ResNet model for visual recognition," *Pattern Recognit.*, vol. 90, pp. 119–133, Jun. 2019.

[37] S. Zagoruyko and N. Komodakis, "Wide residual networks," 2016, *arXiv:1605.07146*.

[38] J. Yu *et al.*, "Wide activation for efficient and accurate image super-resolution," 2018, *arXiv:1808.08718*.

[39] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Nov. 2016, pp. 1637–1645.

[40] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2015.

[41] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 3147–3155.

[42] X. Ding, Z. Lin, F. He, Y. Wang, and Y. Huang, "A deeply-recursive convolutional network for crowd counting," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 1942–1946.

[43] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. ICML*, vol. 30, no. 1, 2013, p. 3.

[44] D. A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," *Comput. Sci.*, vol. 1511, no. 07289, pp. 1–14, Feb. 2015.

[45] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. Eur. Conf. Comput. Vis.* Amsterdam, The Netherlands: Springer, 2016, pp. 391–407.

[46] H. Zhao, X. Kong, J. He, Y. Qiao, and C. Dong, "Efficient image super-resolution using pixel attention," in *Proc. Eur. Conf. Comput. Vis.* Glasgow, U.K.: Springer, 2020, pp. 56–72.

[47] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6848–6856.

[48] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "GhostNet: More features from cheap operations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Nov. 2020, pp. 1580–1589.

[49] S. W. Zamir *et al.*, "Learning enriched features for real image restoration and enhancement," in *Computer Vision*. Glasgow, U.K.: Springer, Aug. 2020, pp. 492–511.

Chang Liu received the B.S. degree in electronic engineering from Xidian University, Shaanxi, China, in 2018, where he is currently pursuing the Ph.D. degree with the School of Artificial Intelligence. His research interests include image soft-decoding, image style transform, and SLAM.

Mingming Ma received the B.S. degree in physics and optoelectronic engineering from Xidian University, Shaanxi, China, in 2017, where he is currently pursuing the Ph.D. degree with the School of Artificial Intelligence. His research interests include hyperspectral image processing, remote sensing, and deep learning.

Fu Li was born in 1982. He received the B.S. and Ph.D. degrees from Xidian University, Xi'an, China, in 2004 and 2010, respectively. From 2010 to 2011, he was with Microsoft Research Asia as a Visiting Scholar. He is currently a Professor with Xidian University. His research interests include brain–computer interface, affective computing, depth sensing, and the application of deep learning.

Zhiwen Chen (Member, IEEE), photograph and biography not available at the time of publication.

Yi Niu (Member, IEEE) received the B.S. and Ph.D. degrees in electronic engineering from Xidian University, Xi'an, China, in 2005 and 2012, respectively. From 2009 to 2012, he was a Visiting Student with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, where he was a Postdoctoral Fellow, from 2012 to 2013. He is currently an Associate Professor with the School of Electronic Engineering, Xidian University.

Guangming Shi (Fellow, IEEE) received the B.S. degree in automatic control, the M.S. degree in computer control, and the Ph.D. degree in electronic information technology from Xidian University, Xi'an, China, in 1985, 1988, and 2002, respectively. From 1994 to 1996, he was a Research Assistant with the Department of Electronic Engineering, The University of Hong Kong, Hong Kong. In 1988, he joined the School of Electronic Engineering, Xidian University, where he has been a Professor since 2003 and is currently the Deputy Director and the Academic Leader of the circuits and systems. Since 2004, he has been the Head of the National Instruction Base of Electrician and Electronic, Xi'an. His research interests include compressed sensing, depth sensing, multirate filter banks, image denoising, image/video coding, and the implementation of algorithms for intelligent signal processing (using DSP and fieldprogrammable gate array).